

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53312 A1

(51) International Patent Classification⁷: C07H 21/04,
C12N 15/11, 15/63, 15/70, 15/82, 15/85, C07K 14/00

(21) International Application Number: PCT/US00/34263

(22) International Filing Date:
26 December 2000 (26.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/488,725 21 January 2000 (21.01.2000) US
09/552,317 25 April 2000 (25.04.2000) US
09/598,042 9 July 2000 (09.07.2000) US
09/620,312 19 July 2000 (19.07.2000) US
09/653,450 3 August 2000 (03.08.2000) US
09/662,191 14 September 2000 (14.09.2000) US
09/693,036 19 October 2000 (19.10.2000) US
09/727,344 29 November 2000 (29.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/488,725 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	09/552,317 (CIP)
Filed on	25 April 2000 (25.04.2000)
US	09/598,042 (CIP)
Filed on	9 July 2000 (09.07.2000)
US	09/620,312 (CIP)
Filed on	19 July 2000 (19.07.2000)
US	09/653,450 (CIP)
Filed on	3 August 2000 (03.08.2000)
US	09/662,191 (CIP)
Filed on	14 September 2000 (14.09.2000)
US	09/693,036 (CIP)
Filed on	19 October 2000 (19.10.2000)
US	09/727,344 (CIP)
Filed on	29 November 2000 (29.11.2000)

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14,

San Jose, CA 95117 (US). ASUNDI, Vinod [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). CHEN, Rui-hong [US/US]; 1031 Flying Fish Street, Foster City, CA 94404 (US). MA, Yunqing [CN/US]; 280 W. California Avenue #206, Sunnyvale, CA 94086 (US). QIAN, Xiaohong, B. [CN/US]; 3662 Tumble Way, San Jose, CA 95132 (US). REN, Feiyan [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). WANG, Dunrui [CN/US]; 932 La Palma, Milpitas, CA 95035 (US). WANG, Jian-Rui [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). WANG, Zhiwei [CN/US]; 836 Alturas Avenue, B36, Sunnyvale, CA 94085 (US). WEHRMAN, Tom [US/US]; 3210 CCSR Mol Pharm, 269 W. Campus Drive, Stanford, CA 94305 (US). XU, Chongjun [CN/US]; 4918 Manitoba Drive, San Jose, CA 95130 (US). XUE, Aidong, J. [CN/US]; 1621 S. Mary Avenue, Sunnyvale, CA 94087 (US). YANG, Yonghong [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). ZHANG, Jie [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). ZHAO, Qing, A. [CN/US]; 1556 Kooser Road, San Jose, CA 95118 (US). ZHOU, Ping [CN/US]; 1461 Japaul Lane, San Jose, CA 95132 (US). GOODRICH, Ryle [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). DRMANAC, Radoje, T. [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

[Continued on next page]

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

WO 01/53312 A1

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such 5 polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2. BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as 10 lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent 15 "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the 20 case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for 25 example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel 30 isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression 35 vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, 5 diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-1786 and 3573-5358. The polypeptides sequences are designated SEQ ID NO: 2n (wherein n = 1 to 20). The nucleic acids and polypeptides are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in 10 the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO:1-1786 and 3573-5358 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a 15 specific domain or truncation of the peptides encoded by SEQ ID NO:1-1786 and 3573-5358 . A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO:1-1786 and 3573-5358 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information 20 from the nucleic acid sequences of SEQ ID NO:1-1786 and 3573-5358 . The sequence information can be a segment of any one of SEQ ID NO:1-1786 and 3573-5358 that uniquely identifies or represents the sequence information of SEQ ID NO:1-1786 and 3573-5358.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on 25 a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid 30 sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing

full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO:1-1786 and 3573-5358 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO:1-1786 and 3573-5358 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., *Science* 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

10 The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO:1-1786 and 3573-5358; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO:1-1786 and 3573-5358; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO:1-1786 and 3573-5358. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO:1-1786 and 3573-5358; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

15 The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO:1-1786 and 3573-5358; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention.

Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of
5 the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the
10 protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA
15 or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as
20 expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide
25 of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition
30 which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein
35 expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides

5 a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the

10 invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a

25 compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound binds to a polypeptide of the invention is

30 identified.

The methods of the invention also provides methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that

35 modulate the overall activity of the target gene products. Compounds and other substances can

effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

10

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule.

Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady

and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells 5 are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

10 As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

15 The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the 20 sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic 25 acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, 30 more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, 35 preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30

nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-20.

5 Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may 10 be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their 15 entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO:1-1786 and 3573-5358. The sequence information can be a segment of any one of SEQ ID NO:1-1786 and 3573-5358 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO:1-20 1786 and 3573-5358. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of 25 human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

30 Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1:4^{25}$) times the increased probability for mismatch at each nucleotide position (3×25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be 35 detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e.g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, i.e., conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations

can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited

5 for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological

10 macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

15 The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (*e.g.*, nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or

20 polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

30 The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3)

35 appropriate transcription initiation and termination sequences. Structural units intended for use

in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed 5 recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked 10 to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells 15 can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (e.g., soluble proteins) or partially (e.g., receptors) from the cell in which they are expressed. 20 "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 25 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

30 The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are 35 described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

5 As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 10 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a 15 listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid 20 sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most 25 preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun 30 Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

35 The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The

term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides 5 which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is 10 determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

15 **4.2 NUCLEIC ACIDS OF THE INVENTION**

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO:1-1786 and 3573-5358 ; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO:1787-3572 and 5359-7144; and a polynucleotide 20 comprising the nucleotide sequence encoding the mature protein coding sequence of the polypeptides of any one of SEQ ID NO:1787-3572 and 5359-7144. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO:1-1786 and 3573-5358 ; (b) nucleotide sequences encoding any one of the amino acid sequences 25 set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO:1787-3572 and 5359-7144. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in 30 receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

5 The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can
10 be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO:1-1786 and 3573-5358 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO:1-1786 and 3573-5358 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO:1-1786 and 3573-5358 may be used as the
15 basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information,
20 representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, e.g., at least about 65%, at least about 70%, at least about
25 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO:1-1786 and 3573-5358, or complements thereof, which fragment is greater than
30 about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in

the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided SEQ ID NO:1-1786 and 3573-5358, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO:1-1786 and 3573-5358 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NO:1-1786 and 3573-5358, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. *J Mol. Evol.* 36:290-300 (1993) and Altschul S.F. et al. *J. Mol. Biol.* 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, e.g., by substituting first with conservative choices (e.g.,

hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid

5 insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and
10 sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the
15 site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids.

20 When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this
25 gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., *supra*, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic
30 code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the 5 polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature 10 protein coding sequences corresponding to any one of SEQ ID NO:1-1786 and 3573-5358, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other 15 nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the 20 invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular 25 organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO:1-1786 and 3573-5358 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into 30 which a nucleic acid having any of the nucleotide sequences of SEQ ID NO:1-1786 and 3573-5358 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially 35 available for generating the recombinant constructs of the present invention. The following

vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

5 The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in*
10 *Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol
15 transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.
20 Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid
25 phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired
30 characteristics, e.g., stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the
35 vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for

transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use
5 can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural
10 sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

15 Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid
20 sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that
25 are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1-1786 and 3573-5358, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic
30 acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO:1787-3572 and 5359-7144 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO:1-1786 and 3573-5358 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a 5 "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO:1-1786 and 3573-5358, antisense nucleic acids of the invention can be designed according 10 to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of a mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of a mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the 15 translation start site of a mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified 20 nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 25 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 30 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiacytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a 35 nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the

inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or

5 genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense

10 nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface

15 receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an

20 α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme.

30 Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave a mRNA transcripts to thereby inhibit translation of a mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO:1-35 1786 and 3573-5358). For example, a derivative of a Tetrahymena L-19 IVS RNA can be

constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SECX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, SECX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

5 Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and 10 Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using 15 standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; 20 Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. 25 PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

30 In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition 35 enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA

portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-*(4-methoxytrityl)amino*-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' 10 DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-1124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the 15 cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 20 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

25 The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association 30 with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (*e.g.*, by homologous 35 recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express

the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter 5 DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotate) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

10 The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the 15 polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, 20 COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using 25 RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant 30 protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived 35 from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK,

HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, 5 SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein.

10 Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast 15 or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial 20 strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of 25 inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, 30 negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the

protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or 5 enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific 10 enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by 15 the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively 20 selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. 25 PCT/US90/06436 (WO91/06667) by Skoultschi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide 30 comprising: the amino acid sequences set forth as any one of SEQ ID NO:1787-3572 and 5359-7144 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO:1-1786 and 3573-5358 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID 35 NO:1-1786 and 3573-5358 or (b) polynucleotides encoding any one of the amino acid sequences

set forth as SEQ ID NO:1787-3572 and 5359-7144 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO:1787-3572 and 5359-7144 or the corresponding 5 full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides 10 comprising SEQ ID NO:1787-3572 and 5359-7144.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., *Bio/Technology* 10, 773-778 (1992) and in R. S. McDowell, et al., *J. Amer. 15 Chem. Soc.* 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, 20 without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also 25 provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid 30 fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, e.g., Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that

retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO:1787-3572 and 5359-7144.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological

methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing 5 an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBatTM kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present 10 invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification 15 of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearlTM or Cibacrom blue 3GA SepharoseTM; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

20 Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, 25 respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG[®]") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP- HPLC) 30 steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

15

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990)).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to

another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprises one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*.
The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers.

Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for

example, Ausubel et al. (eds.) *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, 10 Anderson, *Nature*, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, *Science*, 244: 1275-1281 (1989); Verma, *Scientific American*: 68-84 (1990); and Miller, *Nature*, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or 15 artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease 20 states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be 30 inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in

the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by 5 homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT 10 International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotate) and/or 15 intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to 20 replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or 25 protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene 30 under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally 35 occurring elements. Here, the naturally occurring sequences are deleted and new sequences are

added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the

5 property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial

10 xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or 20 inactivated in the germ line of animals using homologous recombination [Capechi, *Science* 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be 25 prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT 30 Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even 35 replacing the homologous promoter to provide for increased protein expression. The homologous

promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express 5 polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or 10 inactivated in the germ line of animals using homologous recombination [Capecchi, *Science* 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be 15 prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT 20 Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the 25 homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or 30 more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The 35 mechanism underlying the particular condition or pathology will dictate whether the

polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or

5 polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or

10 indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

15 The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as

20 an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of

25

30

35

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

5 Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

10 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch 15 and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

20 Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the 25 polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

30 A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one 35 or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient

confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK,

5 HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., *J. Immunol.* 137:3494-3500, 1986; Bertagnolli et al., *J. Immunol.* 145:1706-1712, 1990; Bertagnolli et al., *Cellular Immunology* 133:327-341, 1991; Bertagnolli, et al., *I. Immunol.* 149:3778-3783, 1992; Bowman et al., *I. Immunol.* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation,

15 Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- γ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells 20 include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., *J. Exp. Med.* 173:1205-1211, 1991; Moreau et al., *Nature* 336:690-692, 1988; Greenberger et al., *Proc. Natl. Acad. Sci. U.S.A.* 80:2931-2938, 1983; Measurement of mouse 25 and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., *Proc. Natl. Acad. Sci. U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 30 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in

35 Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober,

Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder

layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce 5 autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and 10 identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be 15 used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, 20 the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated 25 cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., 30 Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention 35 exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell

sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells 5 proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis 10 and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy 15 to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of 20 various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as 25 those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

30 Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular 35 and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., 5 Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, 10 Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

15

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

20

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of 25 artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

30

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular

endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

5 A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

10 Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

15 Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

20 **4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY**

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including 25 severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be 30 treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

35 Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus,

rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., *Toxicology* 125: 59-66, 1998), skin prick test (Hoffmann et al., *Allergy* 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., *Arch. Toxocol.* 73: 501-9), and murine local lymph node assay (Kimber et al., *J. Toxicol. Environ. Health* 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic

composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration 5 of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in 10 rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic 15 compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may 20 reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating 25 autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 30 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral 35 infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfet them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production,

5 Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E.

10 M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 25:13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population.

Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

5 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. 10 M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

15

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including 20 hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

25

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

30

4.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For 35 example, the presence or increased expression of a polynucleotide/polypeptide of the invention

may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

5 Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic
10 cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps
15 associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central
20 nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

25 Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

30 The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination
35 with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide,

Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Flouxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide,

5 Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin,

10 Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

15 *In vitro* models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wiley-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., 20 Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cell lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, 30 receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, 35 integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen

recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

5 The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. 10 Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

15 By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIACore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a 20 partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and 25 carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.10.13 DRUG SCREENING

30 This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant 35 nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such

transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, 5 which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules. 10

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for 15 screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or 20 organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see 25 Al-Obeidi et al., *Mol. Biotechnol.*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol.*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits 30 modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or 35 cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding

molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

5

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used 10 to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. 15 Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the 20 polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules. 25 The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating 30 the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

30

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or

disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral

5 nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

10 (ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

15 (iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

20 (v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

25 (vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

30 (viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

35 Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or

differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- 5 (iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, *J. Neurosci.* 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, *Exp. Neurol.* 70:65-82) or Brown et al. (1981, *Ann. Rev. Neurosci.* 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by 15 assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as 20 well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy 25 (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, 30 including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female 35 subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or

elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen 5 in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

10

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such 15 polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a 20 polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally 25 involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a 30 single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the 35 present invention can be used to detect polymorphisms. The array can comprise modified

nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could
5 also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid
10 arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a
15 suspension of killed *Mycobacterium tuberculosis* in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed *Mycobacterium tuberculosis* in CFA followed by immediately administering the
20 test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of *Mycobacterium* CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

25

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications
30 include, but are not limited to, those exemplified herein.

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or
35 disorder that can be modulated by regulating the peptides of the invention. While the mode of

administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, 5 condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01 μ g/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1 μ g/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art 10 and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

15 **4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF
ADMINISTRATION**

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered 20 to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the 25 effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell 30 factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic

factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

5

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous,

10 intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

15 Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery 20 system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

25 The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

30

4.12.2 COMPOSITIONS/FORMULATIONS

35 Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be

manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present

5 invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or
10 other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol.
15 When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers

enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding 5 suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents 10 may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be 15 added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as 20 lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of 25 tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or 30 other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for 35 injection may be presented in unit dosage form, e.g., in ampules or in multi-dose containers, with

an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well

known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent.

5 Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

10 The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically

15 acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

20 The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following 25 presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as 30 well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as 35 micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable

lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated 5 herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active 10 ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the 15 various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μ g to about 100 mg (preferably about 0.1 μ g to about 10 mg, more preferably about 0.1 μ g to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition 20 topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other 25 active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or 30 cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the 35 compositions will define the appropriate formulation. Potential matrices for the compositions

may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential

5 matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and

10 biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

15 A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the

20 protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), and

25 insulin-like growth factor (IGF).

30 The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which

modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and

5 with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

10 Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of 15 proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include 20 compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in 25 the method of the invention, the therapeutically effective dose can be estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of 30 the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell 35 cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the

population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD₅₀ and ED₅₀. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 µg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

30 4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the

invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

5 Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , $F_{ab'}$ and $F_{(ab')2}$
10 fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes,
15 subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 1787, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope.
20 Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.
25

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of a related protein that is located on the surface of the protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte
30

Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety.

Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, 5 fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or 10 monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

15 5.13.1 Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic 20 protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an 25 adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of 30 adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG 35 fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the

target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

5

5.13.2 Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro. The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HPRT or HGPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego,

California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or 10 enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

15 After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown *in vivo* as ascites in a mammal.

20 The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the 25 invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or 30 myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the 35 coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin

polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

5 **5.13.2 Humanized Antibodies**

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, 10 immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., 15 Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the 20 humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human 25 immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

5 **5.13.3 Human Antibodies**

Fully human antibodies relate to antibody molecules in which essentially the entire 30 sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL 35 ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal

antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

5 In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon
10 challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al., (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and
15 Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The
20 endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate
25 transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a
30 polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the 5 locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in 10 U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

15 In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

20 5.13.4 F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of 25 monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the 30 treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

5.13.5 Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the

binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan *et al.*, *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to

stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

5 Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment 10 was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity 15 of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from 15 recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region 20 to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, 25 the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific 30 antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991). Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for 35 IgG (Fc γ R), such as Fc γ RI (CD64), Fc γ RII (CD32) and Fc γ RIII (CD16) so as to focus cellular

defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest 5 binds the protein antigen described herein and further binds tissue factor (TF).

5.13.6 Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies 10 have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins 15 can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

5.13.7 Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as 20 to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) 25 and Shope, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

30

5.13.8 Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a 35 cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin,

5 Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), *momordica charantia* inhibitor, curcin, crotin, saponaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the trichothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

10 Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridylthiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as 15 bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., *Science*, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See 20 WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn 25 conjugated to a cytotoxic agent.

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to 30 any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can 35 be used to create a manufacture comprising computer readable medium having recorded thereon

a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

5 A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer
10 readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded
15 thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO:1-1786 and 3573-5358 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO:1-1786 and 3573-5358 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer
20 software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may
25 be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the
30 present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored
35 therein a nucleotide sequence of the present invention and the necessary hardware means and

software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

5 As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially 10 available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present 15 computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide 20 residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a 25 three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

30

4.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA.

35 Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are

designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 152:41:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca

5 Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems.

Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

10

4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

15 In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also

20 comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

25 In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

30 In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

35 Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid

probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of 5 a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

10 Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO:1-1786 and 3573-5358, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

15 (a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of 20 the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the 25 invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time 30 sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the 35 activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to

activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the 5 invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques. 10

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can 15 readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspaczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly 20 described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or 25 multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

30 Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca 35 Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription

from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems.

Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

5 Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

10 **4.19 USE OF NUCLEIC ACIDS AS PROBES**

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO:1-1786 and 3573-5358. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NO:1-1786 and 3573-5358 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, *in situ* hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides 20 additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the 25 cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The 30 nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include *in situ* hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent *in situ* hybridization of

chromosome spreads has been described, among other places, in Verma et al (1988) *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples 5 of genetic map data can be found in the 1994 Genome Issue of *Science* (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

10 4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of 15 skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) *J. Clin. Microbiol.* 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) *Mol. Cell Probes* 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all 20 references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin 25 interaction as a linker. For example, Broude *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. 30 Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell 30 surface termed Covalink NH. Covalink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. Covalink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to Covalink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) *Anal. Biochem.* 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the

5 CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

10 More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm₇), is then added to a final concentration of 10 mM 1-MeIm₇. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

15 Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm₇, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

20 It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard 25 conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

30 An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor et al. (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness et al. (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the 5 light-generated synthesis described by Pease *et al.*, (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 10 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, 15 including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples 20 may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

25 Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA 30 fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation

of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a 25 nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm², depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the 30 subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane.

Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm² and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid 5 being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations 10 may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and 15 variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

20 5.0 EXAMPLES

5.1.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome 25 using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

30 In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

5.1.2 EXAMPLE 2

Assemblage of Novel Nucleic Acids

The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 3573-5358 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend 5 the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage 10 with BLAST score greater than 300 and percent identity greater than 95%.

A polypeptide was predicted to be encoded by each of SEQ ID NO:3573-5358 as set forth below. The polypeptides was predicted using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptides based on a comparison of translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 15 (1990), herein incorporated by reference. The predicted polypeptides are shown in Table 7.

5.2.2 EXAMPLE 3

Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame 20 shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank. Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS:1- 327.

25 Table 1 shows the various tissue sources of SEQ ID NO: 1-327.

The nearest neighbor results for SEQ ID NO: 1-327 were obtained by a FASTA version 3 search against Genpept release 117, using FASTXY algorithm. FASTXY is an improved 30 version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for SEQ ID NO: 1-327 from Genpept . The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The nearest neighbor results for SEQ ID NO: 1-327 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the

signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were

5 examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from

10 Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference.

15 A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.3.2 EXAMPLE 4

20 Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 117, gb pri 117,

25 UniGene version 117, Genpept release 117). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 328-1413.

Table 1 shows the various tissue sources of SEQ ID NO: 328-1413.

30 The nearest neighbor results for SEQ ID NO: 328-1413 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 118, using BLAST algorithm. The nearest neighbor result showed the closest homologue for SEQ ID NO: 328-1413 from Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in

the Sequence Listing. The nearest neighbor results for SEQ ID NO: 328-1413 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., *J. Comp. Biol.*, Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were 5 examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., *Nucleic Acids Res.*, Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were 10 examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from 15 Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" *Protein Engineering*, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by 20 reference. A maximum S score and a mean S score, as described in the Nielson et al as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

25 **5.3.2 EXAMPLE 5**

Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was 30 checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 117, gb pri 117, UniGene version 117, Genpept release 117). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1414-1652.

Table 1 shows the various tissue sources of SEQ ID NO: 1414-1652.

The nearest neighbor results for SEQ ID NO: 1414-1652 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 118, using BLAST algorithm. The nearest neighbor result showed the closest homologue for SEQ ID NO: 1414-1652 from

5 Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The nearest neighbor results for SEQ ID NO: 1414-1652 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were 10 examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were 15 examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determined from using Neural Network SignalP V1.1 program (from 20 Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in 25 each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.4.2 EXAMPLE 6

30 Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 118, gb pri 118,

UniGene version 118, Genpept release 118). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1653-1745.

5 Table 1 shows the various tissue sources of SEQ ID NO: 1653-1745.

The homology for SEQ ID NO: 1653-1745 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 118, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1653-1745 from Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologues 10 with identifiable functions for SEQ ID NO: 1653-1745 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, 15 the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain 20 within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also 25 disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al as reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in 30 each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.5.2 EXAMPLE 7

Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 119, gb pri 119, UniGene version 119, Genpept release 119). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1746-1768.

Table 1 shows the various tissue sources of SEQ ID NO: 1746-1768.

10 The homology for SEQ ID NO: 1746-1768 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 119, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1746-1768 from Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologues with identifiable functions for SEQ ID NO: 1746-1768 are shown in Table 2 below.

15 Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

20 Using the PFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the PFam score for the identified domain within the sequence.

25 The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "30 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et al. reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.6.2 EXAMPLE 8**Novel Nucleic Acids**

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 120, gb pri 120, UniGene version 120, Genpept release 120). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 1769-1786.

Table 1 shows the various tissue sources of SEQ ID NO: 1769-1786.

The homology for SEQ ID NO: 1769-1786 were obtained by a BLASTP version 2.0al 15 19MP-WashU search against Genpept release 120 and the amino acid version of Geneseq released on October 26, 2000, using BLAST algorithm. The results showed homologues for SEQ ID NO: 1769-1786 from Genpept. The homologues with identifiable functions for SEQ ID NO: 1769-1786 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. 20 Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) 25 pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and 30 their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their 35 cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by

reference. A maximum S score and a mean S score, as described in the Nielson et al reference, was obtained for the polypeptide sequences. Table 5 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5

Table 6 is a correlation table of all of the sequences and the SEQ ID NOS.

TABLE 1

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
adult brain	GIBCO	AB3001	9 19-21 50-51 65-66 72 78 80 82 85 87 107-108 113 116 123 138 140 150-152 159 169 177 192-193 202-203 212-214 225-226 235-236 251 258 268-269 272 280-281 295 298 301 321 326 331-332 334 356- 357 362 369 379 382-383 416 423 443 459-460 473 475 477 488 496 500 503 519 526 547 574 582 587 608-609 613 618 633-634 645-646 652 657-658 660 669-671 678 687 695 697 710 715 724 731 775-777 796 804 811 857-859 862 869 899- 900 912 919 922 924-929 933 936 962 979 988-989 996 1001 1004- 1008 1018 1039 1047 1059 1064 1067 1070 1078 1082 1107 1113 1116-1117 1131 1134-1137 1140 1149 1151 1157 1180 1206 1229 1234 1241 1243 1258 1272-1273 1279 1288-1290 1294 1307-1308 1312 1320 1323 1330 1356 1360- 1361 1368 1373-1375 1379 1391 1400 1417 1446 1468 1482 1493- 1494 1501-1503 1506-1507 1512 1517 1522-1524 1530-1533 1537 1549 1565 1578 1598 1606 1608 1623 1625 1627 1639 1643 1648- 1649 1653 1664 1667 1671 1696 1734 1741 1743-1744 1760-1761 1771
adult brain	GIBCO	ABD003	3 12-14 18-19 25 30-31 34-36 43- 45 50-51 56 58 60 65-66 68-69 80 82 85 87 92 104 107-108 112-113 115-116 123-124 131-132 135-137 139 142 146 148-149 152 154 157 159 163 165 167 169 172 180 192- 193 196-197 199 203 208 210 212- 214 223 233 235-237 247 257 259 261 268-269 272 276 280-281 284- 288 291-292 295 297 300-301 304 307 317 320-321 323 327 329-331 333-334 345-349 356-357 379-381 393 401 408 414 419 424 426-428 430 433-436 438-439 443 445 449 453-454 459-461 468 471-473 476- 478 483 491 494 496 500 503 507- 508 516 519-520 525-527 534 536- 540 542-543 545 553 555 560 569- 570 574-576 586-588 593 595 597 601 606-609 616-620 622-623 625 628-633 635-636 643 645-649 653 655-656 660-665 668-670 676 681 687 701 710 715 717 724-728 735 743 745-746 750 753 759 765-766 773 775-778 786 789 796 799-800 802-803 810-811 815 817 820-821 832 834-836 840 845-847 851 858- 861 864 869 874 878 883 897 901- 902 904-905 908 911-914 916 921- 922 924-927 929 932-934 936-939 941-942 945 955-958 963 966-969 977 979-980 985-986 990 992-993 997-1001 1005-1007 1012 1017- 1020 1023-1024 1029-1031 1034 1036 1039 1050 1059 1063-1066 1078 1081-1082 1085-1086 1089

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1097 1103 1107 1109 1112 1116- 1117 1119 1121 1124 1127 1130 1134 1144-1145 1149 1151 1157- 1158 1167 1170 1178 1184 1188 1190 1193-1194 1200 1202 1215- 1217 1220 1226-1227 1229 1231 1241 1243 1247 1252 1258 1263 1267 1269 1279 1281 1284 1286- 1289 1293-1294 1306-1307 1312 1316-1320 1326 1333 1338 1341 1344 1348 1351 1355-1357 1368 1374 1377 1380 1386 1389-1390 1394 1400 1409 1414 1422-1423 1425-1427 1437 1443 1446 1454 1456 1458-1459 1468 1470-1472 1478 1482-1483 1487-1488 1493 1497 1499 1506 1508-1511 1517 1522-1524 1530-1533 1545-1546 1548-1550 1552 1557-1559 1563 1565 1567 1569 1571 1586 1588 1591 1593 1595 1598-1601 1608 1611 1620-1621 1624-1626 1628 1630-1632 1636 1640-1641 1644- 1645 1647 1649 1653-1655 1657 1664 1667 1669 1673 1678-1681 1686 1690 1694-1696 1701 1709 1711 1719 1722-1723 1726-1727 1731-1733 1738 1740 1743-1744 1747 1749 1753 1757-1758 1760- 1761 1765 1771 1785
adult brain	Clontech	ABR001	9 29 68-69 113 115 146 152 206 223 245 277 307 320 324 330-331 344 348 352 362 379 384 393 404 408 414 441-442 454 469 481 490 506 517 586 597 631 641 659 691 715 799 803 833 865 871 875 880 882 908 920 937 1000 1005-1006 1027 1036 1041 1043 1075 1107 1112 1121 1127 1136-1137 1144- 1147 1231 1238-1239 1280 1293 1320 1345 1355 1361 1383-1384 1400 1417 1448 1456 1476 1507 1570 1572 1609-1610 1614 1620 1626 1645 1653 1754 1759 1770 1786
adult brain	Clontech	ABR006	5-8 15-16 168 212-213 271 278 280-281 291-292 300-301 310 314 321 326 336-338 341 352 357 359- 360 362 369 374 379 384 393 396- 397 414 419-420 426-428 430 441- 442 453 506 616-617 661 689 785 798 845 1018 1109 1113 1124 1148 1167 1187 1207 1227 1262 1265 1285 1312 1317-1319 1324-1327 1344 1369 1381 1400 1416 1421 1427 1430-1431 1436 1471 1501 1557-1559 1586 1588 1651 1653 1664-1665 1671 1673 1690 1697- 1698 1700 1711 1717 1719-1720 1728 1736 1740 1743-1744 1757 1760-1761
adult brain	Clontech	ABR008	5-10 13-19 22-23 25 29 33 37-39 43-45 50-51 54-55 57-58 60-66 68-70 72 75 77-80 83 85 89-92 94 99-105 108-110 112-113 116-117 123 128 133 135-137 139 143 145- 146 148 152 154-155 157 166 168- 172 174-175 181-184 188-190 193- 194 196 198-200 202 204-205 207-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			208 210 214-215 218 221-226 229 231-232 234-241 245-247 251-253 255 257-259 268-269 271 276-281 285-286 288 290-292 300-302 304 307 309-311 313 315 317-318 320- 322 325-326 328 330-331 333-338 341 344-347 349 352 354 356-357 362 369-373 376 379-380 382 384 387 390-391 393-394 397 399-403 405-411 414-415 417-420 426-428 437-438 440-444 453-455 462 464 467 469-471 476 478 482-484 488- 491 497 503 506-513 516-517 520 524-526 528-530 532-534 537-540 542 544 547-551 553 561 565-567 572-574 577 581 585 587-588 590- 591 597 599 601-602 606-610 612 615-617 619-620 622-623 628-629 631 633-634 636-641 643 645-647 651-653 655-664 669-671 673 679 682 687 689 691-700 702 706 710 715-717 720-721 725-734 736-739 742-743 746 750-752 756 758-759 762-764 766 768 773-778 780-782 784-785 787-789 794 796 799 802- 803 805 811 814-815 818 825-826 834-837 839-840 842-843 856-859 861-862 865 867-872 874-875 881 883-884 887 889-892 894-895 897- 898 901 904 908 910 912 914 917 919 921-924 926-927 930-932 935- 941 943 945 949 953-954 958 961- 963 967 969 971 975 977 981-983 986 988-990 992 997 999-1002 1004-1006 1008 1012 1018-1023 1027 1029-1031 1035-1037 1047- 1048 1053 1057 1059 1063 1068 1070 1072-1075 1077 1081-1083 1085-1093 1095-1096 1108-1112 1114-1125 1127 1131-1133 1135- 1138 1142-1145 1148-1158 1160- 1163 1167 1169 1172 1175 1177 1180 1183-1188 1191-1195 1199- 1200 1204 1206 1211 1213-1216 1222-1223 1226-1227 1229-1231 1234-1235 1241-1242 1244-1263 1266 1269-1271 1276-1277 1279- 1281 1284-1286 1292 1294-1295 1299 1305-1309 1312 1314 1316- 1319 1322 1324-1327 1330 1332 1334-1335 1339 1344-1346 1351 1354-1355 1357-1358 1365-1367 1369-1370 1373-1374 1376-1379 1381-1384 1386-1388 1392 1394 1396-1397 1400 1403-1407 1410 1414 1419-1420 1423 1432-1433 1435 1437-1438 1440-1442 1446 1448 1453-1455 1457 1461 1463- 1464 1466 1468 1471 1477 1480 1482-1483 1496 1502-1504 1507- 1509 1513 1519-1520 1524-1526 1536 1547 1549-1552 1567 1573- 1574 1578 1586-1589 1597-1598 1601-1602 1605 1607-1609 1611- 1617 1619-1621 1623 1625-1626 1635-1641 1643-1645 1649 1651 1653 1656-1658 1664 1669 1671- 1674 1676-1684 1686 1689-1690 1694-1696 1704-1705 1708-1709

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1720-1724 1726-1728 1730-1733 1737-1740 1742-1745 1753 1756- 1757 1759-1761 1765 1767 1771- 1772 1776-1777 1779-1780 1786
adult brain	Clontech	ABR011	24 75 103 186 210 310-311 364- 365 508 623 710 937 1002-1003 1059 1204 1609 1731-1732
adult brain	BioChain	ABR012	46 182-184 204-205 300 739 767 1371 1549 1620 1684
adult brain	Invitrogen	ABR013	185 204-205 364-365 393 497 595 687 692-694 830 845 1068 1320 1413 1640
adult brain	Invitrogen	ABR014	187 301 357 364-365 375 454 463 731 859 939 983 1073 1262 1270 1320 1403 1640 1651 1657 1696 1722 1738
adult brain	Invitrogen	ABR015	419 434-435 441-442 763 789 983 1320
adult brain	Invitrogen	ABR016	312 364-365 379 1320 1334-1335 1674 1722 1785
adult brain	Invitrogen	ABT004	14-16 22-23 25 37-39 43 58 60 70-72 78 86 94 107 113 116 136- 137 143 146 152 161 173 182-184 194 196 198 210 218 229 259 267 295 298 309-310 320-321 324 336- 338 346-347 349-350 356-357 362 371 379-380 382-383 391 393 396 399 401 408 428 438 459 461 476 482 490 502 507-509 516 526 531 557 562 597 602 607-609 624 652 655 667 669 671-672 687-689 695- 696 710 712 715 721 732 739 743 750 753 766 778 780-781 789 803 814 826 830 837 841 857 869 874 894-895 925 937 949 954-956 960- 961 963 968-969 988-989 1000 1005-1006 1016-1019 1021 1036- 1037 1052 1086 1090 1109 1113 1115 1120-1121 1123-1124 1136- 1137 1140 1144-1147 1151 1167 1170 1174 1188 1193-1194 1205 1225 1229 1231 1254 1258 1262 1280 1285 1309 1312 1334-1335 1341 1343-1344 1356-1357 1370 1378-1379 1383-1384 1403-1404 1423 1429 1434 1442 1448 1451- 1452 1454 1470-1472 1482 1499 1525 1528-1529 1532 1536 1547 1554 1557-1559 1561-1562 1567 1585 1588 1590 1595 1601-1604 1608 1610-1613 1615 1619 1624 1627 1640 1644 1647 1660 1664 1666 1670 1675 1696 1704 1715 1723 1727 1738 1760-1761 1768 1779 1785-1786
cultured preadipocytes	Strategene	ADP001	5-8 11 17 25 68-69 80 82 87 103 105 110 116 136-138 168 171 188- 189 196-198 261 267 276 288 293 301 318 331 336-338 379-380 391 400 428 430-431 510-512 520 524 527 549 557 561 602 618 620 622 631 637 647 670 681-682 710 731 748 782 793-794 817 834-836 843 845 858-859 879 882 893-895 934 960 982 986 995-996 1000 1002 1005-1007 1025 1027-1028 1032 1039 1045 1071 1078 1097 1099- 1102 1136-1137 1140 1219-1220

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1260 1271 1297-1298 1314 1320 1322 1329 1339 1345 1365-1366 1370-1371 1398 1408 1423 1431 1437 1466 1468 1533 1539 1594 1602 1608 1614 1631 1649-1650 1660 1662 1673 1687-1688 1696 1711 1719-1720 1742 1746 1749 1760-1761 1765 1767 1771 1785
adrenal gland	Clontech	ADR002	4-10 15-16 25 29-31 43-45 47 50- 51 55 60 62-63 65-66 75 80 102 116 118 122 126 130 137 150 169- 170 181 192 198 201-203 215 227- 228 247 251 255 267-269 271, 280- 281 285 295 298 311 336-338 342 349 351-352 354 372-373 383-385 391 400 410 415-416 424 426-427 431 434-437 439 445 454 461 473 477 483 491 493 497-498 503 516 519 527 535 546 549 552 572-573 581 588 595 600 602 608-610 620 628-630 637 645-646 670 679 703 713 715 719 732 734 744-746 758 773-778 789 816 829 837 845 848 869 875 883 898 904 912 922-923 930-931 942 948 952 965 967 969 976-977 981 990 992-993 1001 1004 1049 1055 1059 1071-1072 1076 1112-1113 1115 1121 1127 1134-1135 1151 1158 1163 1175 1181 1188 1209 1218 1224-1225 1227 1231 1243 1270-1271 1274 1280 1285 1290 1293 1307 1324- 1325 1327 1330 1342-1343 1345 1348 1365-1366 1369 1378-1379 1387 1398 1400 1405 1417 1425- 1426 1436 1440-1441 1444 1454 1463-1464 1488 1491 1507 1512 1538 1546 1567 1573-1575 1588 1598 1609 1614 1618 1622 1624 1627 1634 1636 1649 1651 1658 1671 1674 1678-1679 1691-1692 1703 1717 1727 1731-1732 1737 1765
adult heart	GIBCO	AHR001	4-8 10-11 15-16 18-21 34-39 44- 46 50-52 57-58 60 62-63 71 75 82 85 87 89 94 97 100 103-104 108- 110 112 114 116 118-119 122-123 127 130-132 134 136-138 141-144 147-151 153 163-164 168-171 179 186 192 195 197 199 204-205 212- 215 220 225-226 229-230 232 234- 236 251 257-260 262 265 272 274 277 280-282 285-286 289-292 296 298-301 304 307 309 314 321 324- 325 330 333 336-338 345 349 351- 352 354 358 361 368 370 380 383- 384 387-388 391 393 397 401 406 408-409 411-412 414-416 430-431 433-439 445-446 449 452 454-455 457 459 462 469 472-473 476-480 483-484 487-490 492-493 496-498 503 506 508 510-513 516 519-522 526 534 536-540 542 546 549 553 560-562 574-577 581-582 584 586- 587 589 593 595 597 604-609 611- 612 615-620 622-623 626 632 637 645-652 656-660 665-666 670-672 674-675 683-684 687 692-694 697 701 709 712 715-716 719-720 725-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			726 728 730-732 735 738-739 743- 744 746 751 753 759 761 765 770- 771 775-780 785 788-790 796 802 804 810 812 817 821 826 828 830 837 843 845-847 849-853 857-861 863-864 869 871 875 877-879 881 883 887 890-892 894-895 897-898 901 903 906-907 911-913 915 919 921-925 927-928 933-935 945 958 961-963 967 969-972 975 977-978 980-986 990 992 999-1002 1005- 1007 1010 1016 1019-1020 1022- 1023 1025 1028-1037 1039-1040 1043 1047 1050 1054-1055 1057 1059 1063-1064 1067-1068 1070 1072 1075-1076 1083 1085-1087 1089 1093-1094 1104 1106 1108- 1109 1113 1116-1117 1119 1121 1124 1126 1128 1131-1134 1144- 1145 1148-1149 1151 1158 1167 1169-1170 1175 1177 1192 1196 1199-1200 1202 1206-1208 1211 1216 1218 1222 1227-1229 1232- 1235 1238-1241 1243-1244 1247- 1248 1250 1253-1254 1256-1258 1261 1268 1270-1271 1277 1280- 1282 1287 1292 1298-1299 1306 1308 1317-1321 1324-1325 1330 1332 1334-1337 1339 1344-1345 1349-1350 1354-1356 1359-1360 1365-1366 1369 1371 1374-1375 1378-1380 1383-1384 1389 1397 1400 1403 1409 1417 1423-1426 1437 1439 1442 1444 1446-1447 1450 1453 1468 1470 1473 1479 1481 1488 1490 1501-1504 1519 1521 1524 1528 1530-1534 1536- 1537 1539, 1541-1542 1547 1553 1555 1560 1565 1567-1571 1588 1591 1597-1598 1601-1602 1605 1614-1616 1619-1620 1623-1628 1630-1632 1634 1636 1641 1644- 1645 1647 1649 1652-1655 1659 1662 1667 1673-1674 1680-1681 1684 1686-1688 1704-1705 1709 1711-1712 1717 1724 1726-1727 1731-1733 1737-1738 1741 1743- 1744 1749 1754-1755 1760-1761 1765 1772 1785
adult kidney	GIBCO	AKD001	4-8 10-11 17-21 29-31 35-39 42- 45 50-51 56-58 60-61 64 68-69 75 77 80 82 85 87 92-94 97 100 102- 104 107-108 112 116-117 119 123 127-133 136-137 139-141 143-144 147-154 157 161-163 165-166 169 172 176 178-179 192 194-197 199 201 203-206 209-210 212-213 215- 216 223-228 234-236 238 247 251- 253 257-259 261-262 265-269 271- 272 274 276-277 279-281 284-286 290 293 296 298-299 301-302 304 307 311-313 321 325-326 329-331 333 341 344 348-350 352 356 358- 359 362 364-365 368 370-372 374 376-377 380-382 392 395 398 400- 401 404 407-409 414-415 423-424 430-437 443-444 446 449 451 453- 455 459 461-462 464 467 469 471- 474 476-477 480-481 483 487-488

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			490-491 493 497-505 510-513 516- 520 522 524 526-529 534 537-540 544 547 549 554-556 560 562 564 567 571-576 578 582 586-589 592- 593 598-599 601 604-606 608-613 615-619 621-626 632-634 637-643 645-652 655 660-664 669-672 675 678-679 688 692-695 698 702 711 713 717 719-720 727 731 735-736 738 743 745-746 751 753 755 762- 763 765 771-773 775-778 780 786 788 793 795-796 800 803 805 808 810-812 814-819 821 826 829 832 834-838 842-845 848-855 857-861 864-865 867 869 871 874 876-883 886-887 889-891 893-896 898-900 902 906-908 910-914 918 920 922 925-927 929-935 937 940-942 945 948-949 951 953-958 960-961 963- 964 969-970 972 976-978 982-986 988-990 992-993 995-997 999-1002 1004-1008 1010 1012-1013 1016- 1017 1019-1020 1022 1025-1031 1035 1038-1040 1042 1044 1047 1050 1054-1055 1057-1064 1068 1070-1073 1078 1085-1086 1088- 1089 1092 1094 1097 1099-1102 1107 1109-1112 1116-1119 1121 1123-1125 1132-1135 1140 1142- 1143 1146-1147 1149-1150 1153- 1154 1157 1159 1163 1167 1170 1178-1179 1181 1183 1192 1196- 1200 1202-1204 1206-1211 1216- 1219 1221-1222 1225 1227-1230 1232-1234 1238-1241 1243-1244 1246-1247 1253 1257-1258 1260- 1261 1267-1268 1270 1272-1274 1281 1283 1287-1289 1293-1295 1299 1306 1308 1311-1313 1317- 1320 1323 1329-1330 1334-1335 1339 1341 1349-1350 1353-1357 1359 1367 1369 1373 1375 1378- 1379 1394 1397 1400 1403 1405 1407-1409 1417 1419 1423-1424 1428-1431 1433 1437-1438 1442- 1443 1445-1446 1448-1450 1453- 1454 1459 1461 1465-1468 1474- 1475 1478 1484-1488 1490 1492- 1493 1495 1497-1498 1506-1507 1509 1512 1518 1521-1522 1525 1527-1528 1532-1533 1537 1540- 1541 1547-1550 1552 1556-1559 1561 1565-1566 1568 1571 1575 1578-1579 1583 1586-1587 1589 1591-1592 1594 1598 1600 1603- 1604 1606 1608 1611 1613 1615- 1616 1618-1622 1624-1628 1631- 1632 1634-1636 1638-1639 1641 1644 1646-1649 1653-1656 1662 1664 1666-1667 1670-1671 1676- 1679 1683-1684 1686 1691-1692 1696-1699 1701 1709-1711 1713- 1714 1716-1719 1723-1724 1726- 1727 1733 1737-1738 1741 1743- 1744 1748-1749 1751 1760-1761 1763-1768 1778 1780 1785
adult kidney	Invitrogen	AKT002	20-21 37-39 47 52 57 60 65-66 68-69 80 104 107-108 122 130 133 136-137 140 142-143 149 169 174

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			181 197 227-228 235-236 244 251 261-265 267 280-281 286 290 299 301 304-305 309 312-313 339 341 344-345 349 358 370-372 376 382- 383 387 392 401 414 416 421 430 443 445 449 453-454 472 487-488 504 506 513 516 519 522 528 536- 540 546 554 585 587 594 598 602 607 616-617 626-627 636 643 662- 664 695 709 721 735 743 761 768 775-777 788 796 804 814 827 837- 838 849-850 852-853 869-870 881 890-892 898 903 905-907 914 919 925 927 934 941 949 952 957 960 962 968 970 1000 1008 1029-1030 1044 1052 1055 1063 1067-1068 1073 1085 1099-1102 1107 1110- 1111 1113 1115 1119 1126 1134 1136-1137 1146-1148 1153 1159 1192 1196 1199 1232-1233 1241 1256 1264 1272-1273 1281 1285 1293-1294 1299 1312 1320 1324- 1325 1330 1344 1349 1351 1355- 1356 1369 1378-1379 1403 1414 1419 1428-1429 1436 1446 1458 1463-1464 1467-1468 1470 1477- 1478 1486 1491 1509 1519 1527 1529 1534 1547 1596 1600 1619 1623 1629 1631 1634 1638 1643 1647 1652 1660 1664 1667 1669- 1670 1673 1686 1709 1727 1740 1776
adult lung	GIBCO	ALG001	4-8 14 37-39 44-46 50-51 56 62- 63 75 82 88 93 103-104 113 125 133 140 143 150 152 154 157 162 171-172 174-175 190-191 196 200 211 214 219 223-224 227-228 251- 252 256 265 272 274 280-281 285 310 332 345 351 362 371 381-382 394 408-409 431 436 445 454 459 461 467 469 471 476-477 488 504 513 527 537-540 544 547-548 554 564 583 607 616-617 621 623-624 634 645-646 662-664 670 695 716 719 743-744 763 766 774 789 803 811 814 817 831-832 837-838 845 852-853 858-859 861 866 880 887 901 905 941 954-957 966 971 977 979 981 987 990 992 996 1001 1005-1006 1014 1017 1045 1047 1054 1059 1062 1064 1072 1080 1086-1089 1094 1107 1126 1134 1136-1137 1142 1150 1157 1173 1190 1200 1208 1220 1241 1272- 1273 1280 1282 1295 1306 1320 1331-1332 1353 1374 1379 1383- 1384 1404 1409 1423 1434 1436 1442 1474 1478 1494 1509 1522 1525 1531-1532 1547 1549 1553- 1554 1571 1598 1606 1613 1624 1627-1629 1632 1642 1644 1662 1669 1676-1677 1684 1696 1727 1731-1732 1737-1738 1748-1749 1786
lymph node	Clontech	ALN001	4 24 50-51 82 105 137 153 198 201 223-224 234 268-269 272 280- 281 287 301 312 329 343 382 421 430 433 445 451 461-462 475 481- 482 503 526 529 537-540 546-547

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			621 626 649 679 719 725-726 738 793 803 831 834-836 838 844 857- 858 866 879 905 913 928 963 976 1005-1006 1012 1038 1050 1116- 1117 1151 1199 1204 1226 1243 1265 1274 1324-1325 1339 1353 1374 1377 1440-1441 1447 1504 1549 1600 1618-1619 1631 1641 1644 1653 1687-1688 1691-1692 1741 1771
young liver	GIBCO	ALV001	5-8 11 20-21 46 50-51 58 65-66 75 79 82 93 97 102-103 108 110 116 139 143-144 148-149 171-172 174 187-189 194-195 198 209 214- 215 230 250 258 267-269 280-281 306 309 342 351 356 359 362 372 374 392 394 398 401 407-408 410 414 431 444 455 459 476 478 483 493 510-512 516 520 522 526 536 549 571 574-577 585 592 601-602 607 621-624 628-630 632-633 637 648 660 666-667 678 697-698 700 717 719 728 730 734 738 744-745 766 770 773 779 788 800 808 812 814 841 849-851 871 874 879 887 893 898-900 902-904 906-907 911 919 922 924 934 953 957 963 965 970 984 986 997 1001 1004 1007 1012 1029-1030 1033-1034 1052 1061 1066 1070 1076 1086 1089 1093 1099-1102 1110-1112 1116- 1117 1119 1121 1125 1136-1137 1144-1145 1156-1157 1159 1196 1199-1200 1209 1211 1219-1220 1241 1244 1262 1270 1275 1279 1283 1295 1317-1320 1332 1339 1344 1359 1362-1363 1379 1383- 1384 1403 1415 1430-1431 1437 1450 1467 1475-1476 1483-1484 1494-1495 1498 1505 1512 1516 1518-1519 1526 1529 1547 1550- 1552 1557-1559 1565 1583 1587 1597 1609 1614 1620 1631 1637 1641 1644 1654-1655 1662 1667 1669 1684 1691-1692 1702 1711 1725 1738 1741 1743-1744 1758 1760-1761 1763-1765 1769
adult liver	Invitrogen	ALV002	5-8 17 20-21 32-33 41 55 58 64 75 77 86 89 102 108 117 119 175- 176 198 200 209 231 235-236 250 272 275-276 284 306 316 321 325 333 356 359 374 376 398 401 408 414 428 430 433-435 454 476 494 503-505 517-518 528 534 544 552 561-563 567 578 581 608-609 630 632 637 644 650 661 665 672 702 707 710 721-722 750 753 778 782 794 814 820 826 834-837 847 849- 850 858 861 874 879 893 898 904 911 918 921-922 926 946 948 972 978 986 996 1020 1027 1031 1034 1053 1063 1068 1070 1073 1086 1089 1093 1097 1113 1119 1156 1159 1195 1198-1199 1208 1220 1227 1241 1261 1272-1273 1277 1285 1308 1315 1320 1324-1325 1330 1362-1363 1375 1403 1408- 1409 1415 1431-1432 1435 1467 1469 1482 1504 1524 1542 1547

Tissue Origin	RNA Source	HySeq Library Name	SEQ ID NOS:
			1550 1567 1578 1581 1583 1594 1597 1601-1602 1611-1612 1615 1618-1619 1621 1625 1637 1645 1647 1652 1654-1655 1660 1666 1669-1671 1684 1706 1722 1737- 1738 1742-1744 1760-1761 1763- 1765 1772 1774
adult liver	Clontech	ALV003	29 676 997 1063 1119 1536 1766
adult ovary	Invitrogen	AOV001	1 4-18 20-23 29 35-40 42-48 50- 51 53-58 61-63 65-66 68-69 73-75 77-78 80 82 85 87 89 97 100-101 103-104 106-108 110 113 115 118 122-124 126 128 133-134 136-140 142 145-147 149-157 161 166 168- 170 174 177-178 180 182-186 188- 189 192-203 207 209 211-215 219 221-224 229-230 234 242-243 246- 247 255 258 260-262 265-269 271- 272 274 277-281 284-286 288 290 295 299 301-302 304 307 309-311 313-314 316 321 323-326 330 332- 333 335-338 341 344 349 352-353 356 358 360 362 370-372 376-377 379-384 387 390-392 394 397-398 400 403 408-410 412 414-416 423- 424 426-427 430-435 439 443-446 448-449 451 453-455 462-463 468- 471 473 476-479 481-484 487 489- 494 496-497 499-501 503-505 509- 514 516-517 519-520 522 524 526 528-534 541-544 546-547 549 552 554-555 561-564 566-567 569-570 572-573 575-576 579 581 583 585- 588 590-591 593 595 597 599 601- 605 607-613 615 618-622 624-627 630 632-633 636-640 642 644-647 649-652 654-655 657-665 667-675 677-678 681 683-684 692-695 697- 710 714-721 723 725-727 729 732 734-735 743-746 750-751 753 758 763 765 767 772-773 775-778 780 783-784 786 788 790-791 794-796 800 803 805 809-811 813-815 818- 819 821-824 826 828-829 831-832 837-838 843-850 852-857 859-864 867 869 871-872 874-875 878-883 887-888 890-895 898-910 912-914 916 919-922 924 926-927 929-939 941 943-946 948-951 953 955-958 961-964 966-967 970-979 981-982 985-986 988-990 992 995-997 999- 1001 1004-1009 1011-1013 1016 1019-1020 1024-1025 1029-1031 1033-1035 1037 1039 1041-1047 1050-1051 1054-1060 1062-1064 1067-1070 1072-1073 1075-1076 1078-1079 1085-1086 1089-1090 1094-1096 1098-1103 1106-1108 1112-1117 1119-1120 1123-1127 1131-1135 1142-1143 1146-1149 1153 1156 1158 1163 1165-1166 1169-1171 1173-1175 1177-1178 1180 1183-1185 1190-1191 1195 1197-1200 1202 1205-1214 1217- 1219 1221-1226 1232-1235 1238- 1241 1243-1244 1247 1249 1252- 1254 1256-1258 1262 1265 1267- 1268 1270 1275 1278 1280-1283 1286-1289 1291 1293-1294 1298-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1299 1306 1308 1312 1317-1321 1323 1327 1329-1330 1332-1333 1338-1339 1341 1343-1351 1356 1359 1361 1365-1366 1371-1375 1377-1379 1383-1384 1386 1389 1394 1400 1404 1416-1417 1422- 1427 1429-1431 1435-1436 1439- 1443 1445-1450 1453-1454 1459 1463-1464 1466 1468 1470 1474- 1481 1484-1485 1488 1491 1493- 1494 1496-1498 1501-1504 1506- 1507 1511-1517 1519 1521-1524 1526-1527 1530-1531 1534-1536 1538-1539 1541 1546 1548-1550 1553 1555-1559 1561-1563 1566- 1567 1569-1570 1572 1574-1575 1578 1580-1581 1587-1588 1590- 1591 1595 1597-1598 1600-1606 1609 1611-1621 1623-1630 1634 1636 1638 1641 1643 1645 1647- 1657 1659-1662 1664 1667 1669- 1671 1673-1674 1676-1681 1683- 1690 1699 1702-1707 1710-1711 1713-1714 1716-1719 1723-1724 1726-1728 1731-1733 1735 1737- 1738 1740-1741 1743-1744 1748- 1751 1753 1755-1756 1760-1762 1765 1767-1768 1770-1771 1776 1778-1779 1783-1784 1786
adult placenta	Clontech	APL001	5-8 44-45 90-91 107-108 159 178 311 351 414 476 503 545 574 624 636 719 755 773 860 890-891 924 947 955-956 962 990 992 1002 1045 1202 1320 1369 1628 1686 1713-1714 1743-1744
placenta	Invitrogen	APL002	14-16 26 29 43 60-61 79-80 103 106 116 135 171 177 180 194 196 198 210 216 235-236 272 290 299 309 329 334 339 359 379-380 417 423 430 434-435 448 454 483 490- 491 517 522 631 723 725-726 728 738 746 769 818 843 854-855 857- 858 916 948 953-954 976 988-989 1005-1006 1013 1033 1036 1064 1068 1070 1086 1139 1144-1145 1160 1277 1285 1317-1320 1343 1345 1429 1435 1438 1454 1482 1486 1490 1512 1519 1532 1549 1592-1593 1602 1626 1647 1649 1664 1673 1675 1722 1727 1730 1746 1776
adult spleen	GIBCO	ASP001	3 5-8 12 15-16 19-21 24 29 34-36 44-45 57 60 82-83 87 89 94 98-99 103 106 108 117 119-121 139 141 147 152-153 155 166 169 171 174 178-180 196 198 201-206 209-211 215 219 234 253-254 256 258 264 272 280-281 290 295 302 309 312 325 333 341 349 358 372 382 386- 387 394 406 414 431 434-436 446 448 451 473 481 490-493 500 503 505 517 519 530 534 536-540 547 554 557 574-576 582 592 595 604 611-612 620-621 623 631-632 642 652 659 661 667 671 673-675 684 700 721 728 730 732 738 742-744 746 762 765 774 780 788-789 794 810-811 817 822 830 832 845 848 852-853 858 862 866 874 879 882

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			884 906-908 912 919 921-923 926- 927 934 942 949 957-958 963 977- 978 983 990 992-994 996-997 999 1005-1007 1010 1012 1031 1036 1042-1044 1046 1049 1059 1068 1070 1076 1089-1090 1094 1103 1109 1113 1115 1124 1140 1163 1170 1174 1177 1190 1196 1219- 1220 1226-1227 1229 1236 1241 1246 1258 1269 1271 1274 1295 1301 1320 1322 1330 1334-1335 1339 1349 1351 1353 1359-1360 1364 1369 1374 1386 1397 1413 1417 1434 1436-1437 1439 1468 1474 1477 1480 1485-1487 1498 1512 1522 1525 1544-1549 1553 1560 1567 1591 1600 1631 1636 1651 1654-1655 1658 1662 1670 1674 1678-1679 1684 1686 1700 1727 1733 1738 1740-1741 1760- 1761 1774 1779 1781-1782
testis	GIBCO	ATS001	5-8 10 26 30-31 47 50-51 57 68- 69 82 84-85 97 102 113 119 137 139 150 152 154 156 163 169 174 176-177 192 194 196-197 212-215 227-228 247 255 258 261 282 285 288-289 301 307 311 316 330 334 349 370-372 392 398 410 415 426- 427 430-431 433 437 446 454 461 469 473 477 481-482 493 499 502- 503 513 522 526 547 552-553 563- 564 572-573 575-576 581-582 585 599-602 605 612 615-617 620 631 637 647 649-650 656 660 665 670 674-675 712 719-721 723 728 731 738 744 746 773 780 784 788-789 802 804 809 811 814 826 831 837 843 845 848 859 866 869 877 905 913 916 919 921 926 929 937 950 960 963 971 975 977 981 990 992- 993 1007 1016 1029-1030 1034- 1035 1038-1039 1045 1059-1060 1064 1070 1072-1073 1087 1089 1097 1099-1102 1104 1108 1113 1141 1149 1161-1162 1175 1208- 1209 1222 1227 1229 1231 1235 1238-1239 1243 1253 1285 1287- 1289 1291-1293 1307 1311 1317- 1320 1330 1332 1338 1345 1369 1373-1374 1379 1389 1399-1400 1409 1423-1424 1430 1435-1437 1443 1459 1484 1486 1490 1493 1496-1497 1501 1505 1509-1513 1527 1530-1531 1533 1537 1546 1549 1563 1565 1567 1569 1571 1577 1586 1591 1599 1602 1625 1628 1630-1632 1636 1639 1642 1649 1661-1662 1666-1667 1670 1675 1684 1690 1699 1705 1712 1717 1724 1730 1737-1738 1752 1767 1779
Genomic DNA from BAC 63I18	Research Genetics (CITB BAC Library)	BAC001	686 1352 1412
Genomic DNA from BAC 393I6	Research Genetics (CITB BAC Library)	BAC002	1411-1412

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
Genomic DNA from BAC 393I6	Research Genetics (CITB BAC Library)	BAC003	1352
adult bladder	Invitrogen	BLD001	5-8 17-18 22-23 33 37-39 56-57 80 93 100 120-121 169 201 237 251-252 272 278 311 348 363 382 413 415 424 430 443 483 502 542- 543 562 564 607 616-617 626 635 652 667 671 710 727 755-756 762 773 786 788 837 840 866 893 898 909 918 929 966 977 983 1016 1025 1055 1073 1082 1140 1167 1185 1189 1199 1270 1369 1481 1536 1560 1573 1596 1614 1636- 1637 1649-1650 1654-1655 1658 1669 1671 1690 1719 1727 1731- 1732 1739 1741 1760-1761 1779
bone marrow	Clontech	BMD001	3-8 11 13 18 29-31 33 35-36 40 43-45 47-48 50-51 57 60 65-66 75 80 82 85 88-89 94 100 103 107 110 115 118-119 124-125 133-134 136-137 139-141 146 150 152-153 155 161 163 168-170 172 178-180 187 192-193 197-198 203-205 210- 213 215 217 219 222 224-226 233 235-237 242-244 255 258 260 263- 264 266 273 276 278 283 286 290 295 301-302 307 312-313 321 330 333 339 343 352 357-358 370-371 382 384-385 387 389 394 408 410 412 416 421 424-427 429-431 436- 437 439 441-442 445 447 454-456 461-462 471-472 475 477-479 481- 482 485 488 493 498 500 503-506 513 516 519 523-524 526 530 535- 540 542 544-545 549 555 565 567 569-577 581 583-586 588 593 601 603-604 608-609 613-619 621-622 632-633 636-637 642 649-650 656- 660 666 670 672 674-675 679 683 701 708 716 718-720 731 735-736 740-742 744-745 752 761 765 772- 773 775-778 780 785-786 789-791 796 798 802 810-812 823-824 826 830 832-833 837-838 843-844 848- 855 858-859 866-867 869 878-880 883 890-892 896 903 905 908 912- 914 922-924 927 930-931 937 939- 941 952-953 955-958 963 969 973 976 981 985 987 990 992 995 1000 1002 1005-1007 1013 1016 1025 1028-1031 1033 1035 1037 1039 1042 1044 1047 1050 1053-1054 1059 1061 1063 1066 1070-1071 1079 1106 1110-1113 1115-1117 1124 1126 1134-1135 1142 1144- 1145 1163 1172 1178 1197 1199- 1200 1202 1216-1217 1224 1227- 1228 1240 1246 1254 1261 1266 1270 1278 1281 1285 1287 1290- 1291 1293 1299-1301 1308 1314 1317-1320 1327 1331 1339 1343 1346 1349 1353 1356 1361 1367 1369 1372-1374 1379-1380 1394 1400 1403 1406 1408 1413 1417 1419 1423 1425-1427 1430-1431 1433 1439 1443 1446-1449 1459 1463-1464 1482 1486 1493-1494

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1506 1509 1513 1521-1522 1524 1526 1528 1531 1536-1537 1543 1546 1548-1549 1552 1554-1555 1557-1559 1571-1572 1581 1589- 1592 1597-1600 1609 1614 1621 1626-1628 1630-1632 1634 1636 1638-1639 1641 1646-1647 1651 1653-1655 1661-1662 1676-1681 1684 1686 1690 1702 1707 1711 1713-1714 1717 1720 1722-1723 1727 1737-1738 1740 1758 1767 1772 1781-1782 1785-1786
bone marrow	Clontech	BMD002	11 15-16 19 30-31 35-36 68-69 75 83-84 93 99 103 108-109 118 137 139 169-170 174 177 180 190 193 212-213 219 222 225-226 232 237 255 259 264 273-274 284 286 290- 292 295 301 303-304 307 312-313 316 324 326 330 334-335 348 352- 353 357 360 370-373 384 386-387 397 403-404 414-416 421 425-427 429-430 433-436 440 444 451 454 465-466 472 475 478 491 493 516 520 523 525 531 545 548 552 566 569-570 581 583 590-591 597-598 601 616-617 621 641 650 652 656 659 671 674-675 679 684 710 718- 719 728 734 737-738 742 761 765 774-778 790 811 814 818 830 834- 836 854-855 859 866 869 871 878- 879 884 889 892 904 922-923 932 990 992 998 1001 1004 1016 1036 1042 1048 1051 1054-1055 1058 1088-1089 1106 1112-1114 1155 1157 1192 1200 1223 1227-1228 1236-1237 1260-1261 1282-1283 1285 1287 1295 1314 1317-1321 1324-1327 1330 1333 1341 1343 1347 1350 1353 1355-1357 1367 1369-1370 1373 1377 1379 1381 1383-1384 1394 1397 1400 1406 1413 1417 1425-1427 1438 1442 1446 1459-1460 1470 1493 1505 1521 1536 1546-1549 1560 1573- 1574 1578 1598-1600 1621 1626 1631 1634 1646 1649 1653 1656 1658 1669-1670 1683-1684 1687- 1688 1690-1693 1696 1699 1702 1704 1707-1709 1711 1720 1722- 1723 1725 1727 1729 1731-1733 1738-1740 1743-1746 1752 1755 1760-1761 1767 1777 1781-1782 1786
bone marrow	Clontech	BMD004	73-74 503 922 1036 1711
bone marrow	Clontech	BMD007	95-95 866 1320 1475
adult colon	Invitrogen	CIN001	17 56-58 103 110 117 144 150 171 179 185 188-189 201 204-206 210 218-221 225-226 231 237 251 277 288 310 312 320 333 359 386 388 394 408 420 455 481 485 503 510- 512 590-591 615 635 647-648 665 672 684 697 710 725-726 743 780 786 788 826-827 848-850 854-855 858 866 872 898 918 921-923 953 976 983 993 1005-1006 1017 1020 1025 1027 1054-1055 1063 1068- 1069 1140 1153 1170 1185 1196 1199 1220 1280 1314-1315 1320 1345 1351 1355 1369 1428 1439

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1462-1464 1512 1556 1583 1587 1594 1596 1614 1625-1626 1631 1639 1645 1650 1675-1677 1687- 1688 1701 1713-1714 1724 1740 1765
Mixture of 16 tissues - mRNAs	Various Vendors	CTL016	401 1490 1686
Mixture of 16 tissues - mRNAs*	Various Vendors	CTL021	312 782 1132-1133 1403 1712 1715
adult cervix	BioChain	CVX001	1 4-8 11 13 18-21 25-26 30-31 33 37-39 43 46-47 58 61 64-66 71 73-74 82 85 94 100 103-104 113 118 122 126 130 134 140 147 153- 156 163 170 179 181 186 192 195- 196 198 201-202 218-219 222 229- 231 257 266 276-277 285-286 288 298 301-302 304 307 312-314 324 326 329-330 332 335 342 352 358 362 371-372 376 379 381-382 384 388 398 400 410 414 416 419-420 426-427 430-431 433-436 439 446 448 461-462 464 471-477 479 482- 483 491 493 496 503 506 510-513 516-517 526 530 535 542-544 546- 547 557 561 572-573 575-577 581- 582 585-586 588-589 593-594 600 602 604-605 607-609 612 615-619 623 644 650 654 657-658 662-665 670 672 680 683 691-694 698 706 708-709 711 713 720-721 727 729 731-732 737 745-747 753-754 760 765 771 774-777 780 790 793 796 798 800 803 805 818 826 828 831- 832 834-836 843 847-848 851-855 857-860 864-866 869 871 876 878- 880 882 887 890-891 897 899-902 905-908 912-913 916 918-919 922 927 932 934-938 944 948 955-956 958 963-964 967 969-970 972 976 978-979 983 985 990 992 1000 1005-1007 1016-1017 1024 1027 1033 1036 1038 1045 1047 1053- 1056 1066-1067 1071 1073 1075 1079 1082 1098 1113 1124 1129 1134 1139 1146-1149 1163 1167 1170 1173 1175 1177 1181 1197 1200 1202 1211 1214 1216 1221- 1222 1225 1227 1232-1234 1240- 1241 1243 1258 1264-1265 1268 1270 1279 1287-1290 1308 1310- 1311 1316 1320 1323 1327 1345 1349 1353-1354 1360 1372-1374 1383-1384 1386 1394 1397 1405-

* The 16 tissue-mRNAs and their vendor source, are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) normal adult kidney mRNA (Invitrogen), 3) normal adult liver mRNA (Invitrogen), 4) normal fetal brain mRNA (Invitrogen), 5) normal fetal kidney mRNA (Invitrogen), 6) normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) human bone marrow mRNA (Clontech), 10) human leukemia lymphoblastic mRNA (Clontech), 11) human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1406 1416 1425-1427 1431 1436- 1437 1442 1446 1448 1453 1459 1466 1472 1478 1482 1496 1501- 1503 1506 1512 1522 1527-1528 1531 1533 1541 1547 1569 1571 1585 1589 1597-1598 1600 1608- 1609 1614-1616 1620 1623-1624 1626-1628 1630 1638 1641 1643 1649 1653 1656 1662 1667 1669 1674-1675 1683 1685-1688 1699 1702 1709-1710 1715 1717 1722 1724 1729 1731-1732 1735-1739 1741 1743-1744 1748-1749 1755 1760-1762 1767 1773 1778 1785- 1786
diaphragm	BioChain	DIA002	137 282 289 730 780 986 1409 1478 1599 1614
endothelial cells	Strategene	EDT001	3 5-10 13 15-21 24-26 29 34 37- 39 42 44-45 50-51 53-55 57-58 60-61 65-66 68-69 73-74 77-78 80 82-83 85 87 89 93-96 101-105 108 110 112-114 116 118-122 124 128 133-134 137-142 147-150 152-153 161-163 166-172 176-179 187 190 192 194 196-201 204-207 210 212- 214 220 224 229-230 233 235-236 240-241 251-252 258 261-262 265 267-269 272 276-277 279-281 284- 285 288 290 295-296 301-302 310- 311 313 316 321 325 329 331-333 335 340-342 351-355 360 371 375 380-382 384 387 390 392 397 400 407-408 410 412 414 416 425-427 431 434-436 439 444-445 449 454 463-464 472-475 477-479 486 488- 490 497-498 500-504 510-513 516- 519 522 524 526-528 532-534 536- 540 542-546 548 561-563 566-567 572-576 579 581 585-586 589 593 595 597 599 603 607-612 615-617 620 622 626 630 632-634 638-641 644 647 656-660 662-664 670 673 678 680-682 692-697 707 709-710 712-713 719 730 732 734 736 738 743-746 751 759 768 771 773 775- 778 783 786-789 793 800 803 805- 807 810-811 814 816-818 821-822 824 826 828-829 832 834-838 842- 845 848-850 854-860 862 864 869 871 874 876-879 883 885 887 890- 891 894-895 898-900 903 908 910- 913 916 919-922 924 926-928 930- 935 939 943 948-949 951-954 957 959-961 964 969-970 973 975-978 983-984 988-990 992-993 996-997 1000 1002 1004-1013 1016-1020 1022-1025 1028 1031 1033-1034 1038-1046 1050 1055-1056 1059- 1060 1062-1064 1067-1070 1072- 1074 1076 1078 1082 1086-1087 1089-1090 1093-1097 1099-1103 1107 1109-1113 1116-1117 1124- 1126 1128-1131 1134-1135 1138 1140 1144-1145 1148-1149 1153 1157 1160 1163 1171 1183-1184 1198-1199 1202 1205-1207 1211 1216-1217 1219 1221 1225 1229 1232-1235 1238-1241 1243-1244 1246 1250 1253 1257-1258 1261

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1265-1266 1268 1270-1271 1274- 1277 1280-1283 1285-1286 1288- 1290 1293 1295 1298 1308 1312 1317-1320 1324-1325 1327 1329- 1330 1334-1335 1338 1342-1343 1345-1347 1350 1355-1356 1359 1367 1369 1374 1376 1379 1398 1400 1406 1408 1414 1417 1419 1424-1426 1428-1431 1434-1438 1440-1442 1448 1450 1462-1466 1468 1472 1474 1478 1487-1488 1491-1493 1501-1504 1506 1509 1511 1516 1520-1521 1526 1529 1531 1536-1537 1539-1540 1546- 1547 1549 1552 1555 1557-1559 1561-1565 1568 1571 1575 1578- 1579 1581-1583 1587-1588 1590 1592 1597 1605-1606 1611 1613 1615 1618-1621 1624-1628 1630- 1631 1634 1636 1638 1641 1643- 1650 1652-1659 1664 1666-1667 1669 1671 1675-1681 1683-1688 1696-1698 1703 1711 1715-1716 1719 1722-1723 1726 1731-1733 1736 1739-1741 1743-1744 1749 1755 1760-1761 1765 1767-1768 1771-1773 1776 1779 1783-1786
Genomic clones from the short arm of chromosome 8	Genomic DNA from Genetic Research	EPM001	286 686 1297 1303-1304 1352 1411-1412 1754
esophagus	BioChain	ESO002	131-132 261 289 380 503 860 892 1000 1007 1397
fetal brain	Clontech	FBR001	62-63 89 112 126 194 322 336-338 379 391 411 481 546 563 607 679 710 867 1012 1031 1055 1251 1262 1320 1407 1643 1652 1686 1731- 1732 1746 1765
fetal brain	Clontech	FBR004	68-69 90-91 139 212-213 301 331 362 374 403 436 611 645-646 659 668 670 691 785 805 845 1163 1209 1216 1232-1233 1238-1239 1387 1410 1416 1430 1496 1536 1547 1593
fetal brain	Clontech	FBR006	5-9 25 43 60 62-63 65-66 70 72 80 87 92 101 103 108 114 136 139 149 152-153 157 168 171-172 175 207-208 210 212-213 221-226 237- 238 251-253 266 272 279-281 295 301-302 307 310 317-318 321-324 330 333-334 336-338 346-347 352 357 370 373 377 379-380 382 384 391-392 397 399 402 406-408 410- 411 417 421 424 426-427 430 436- 437 440-443 454 460 464 467 473 476 483 488-489 495 497 508 510- 513 516 519-520 524 530 537-540 544 547 550 561 567 572-574 582 590-591 595 597 604 607-609 615 623 628-629 631 634 638-640 655 657-658 660 665 669 674-675 679 689 691-694 696-697 699 701 706 710 716 720 728 732 734 736 742- 744 757-760 763 775-778 780 799 806-807 810 817-818 826 839 843 858 861 864 871-872 884 890-891 894-895 898 904 915 921-923 935- 936 938 945 950 952 955-956 958- 959 961 963 967 969-971 990 992

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			999 1001 1005-1006 1008 1013 1016 1022 1024 1029-1030 1032 1035 1042 1047-1048 1052 1056 1065 1067 1070 1082 1089 1109 1114-1115 1119 1131 1143-1149 1151 1153-1156 1160 1163 1167 1172-1173 1178 1184 1186 1188 1190-1200 1211 1216 1222-1223 1226-1227 1229 1231 1236 1245 1253-1255 1258 1260 1262 1266 1270-1273 1281 1287 1308-1309 1314 1317-1320 1326 1334-1335 1339 1341 1344 1350 1356 1369- 1371 1373 1376 1379 1381-1382 1386 1392 1396-1398 1419 1423 1425-1426 1428-1429 1432 1437 1440-1441 1448 1466 1470 1482 1502-1503 1507 1511 1513 1516 1519 1536 1544 1549-1550 1557- 1559 1573 1589-1590 1598 1608 1611-1614 1619 1621 1625-1626 1640 1651 1657-1658 1676-1679 1693 1696 1703-1704 1713-1714 1718 1720 1722 1724 1726 1728 1730-1733 1735-1736 1738-1739 1742 1745 1755 1759-1761 1765 1767 1771-1772 1777 1779-1780 1786
fetal brain	Clontech	FBRS03	235-236 520 864 1068 1188 1587
fetal brain	Invitrogen	FBT002	15-18 20-21 24-25 29 34 43 61-63 77-78 98 101 103 107-108 128 130 136 146 148 165-166 171 174 181 185 196-198 204-205 208 223 230 235-236 251 253 261 268-269 280- 281 284-285 288 309-311 321 329 334 339 346-347 350 357-359 381- 383 390 407 418-419 430 434-435 438 443-444 461 464-466 483 490 494 509 516 519 522 527 557 561- 562 572-573 590-591 595 597 623 632 647-648 650 655 669-670 672 682 690-691 700-701 710 717 736 746 782 784 788-789 814-815 825 829 840-841 847 854-855 857-858 897-900 904 919 925 935-937 946 948-949 954 960-962 966 969-970 986 996 1000-1001 1005-1007 1012 1014 1022-1028 1045 1052 1055 1068 1070 1072 1078 1082 1085 1090 1109 1115 1118 1120 1128 1136-1137 1144-1145 1149 1156- 1157 1193-1195 1198 1204-1205 1220 1222 1234 1257 1262 1271 1274-1275 1280 1285-1286 1294 1312 1314 1317-1320 1330 1342 1344-1345 1349-1350 1355-1356 1358 1364 1369 1379 1383-1384 1431 1435 1476 1507 1519 1532 1536 1547 1554 1564 1567 1578 1582 1587 1593 1595 1601 1608 1615 1619-1621 1638 1644 1661 1665-1666 1673 1687-1688 1690 1715 1723 1728 1749 1753 1757 1759-1761 1765 1771 1774 1776 1778 1781-1782 1786
fetal heart	Invitrogen	FHR001	105 124 180 289 864 1036 1148 1229 1614 1616 1762 1785
fetal kidney	Clontech	FKD001	5-8 11 40 47 57 65-66 82 85 102 124 163 171 216 222 224 235-236

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			258 277 280-281 307 310 314 330 371 387 392 395 403 422-423 431 436 443 455 469 500 519 522 542 563 572-573 585 600 619 623 650 654 657-658 660 679 719 731 780 798 821 833 844 854-855 857 864 868 878 911 929 958 960 969 990 992 1007 1046 1087 1103 1129 1139 1285 1312 1331 1355 1369 1371 1376 1391 1422 1425-1426 1440-1441 1470 1543 1598 1601 1618 1631 1651 1654-1655 1669 1678-1679 1691-1692 1733 1785
fetal kidney	Clontech	FKD002	352 384 426-427 440 583 602 1060 1131 1324-1325 1636
fetal kidney	Invitrogen	FKD007	20-21 82 163 335 679 988-989 1000 1227 1230 1320 1554
fetal lung	Clontech	FLG001	35-36 94 323 371 398 426-427 445 473 549 560 604 616-617 626 631 649 651 719 746 786-787 832 842 849-850 864 894-895 1075 1178 1182 1200 1206 1309 1311 1345 1429 1493 1567 1576 1620 1686
fetal lung	Invitrogen	FLG003	9 15-16 29 41 47 68-69 83 88-89 102 124 137 152-153 165 196 224 229 231 249 254 256 267 291-292 300 325 333 344-345 352 373 376 379 384 408 426-427 430 432 467- 468 475 483 488 493 516 531 535 545 547 549 564 582 602 623 644 660 662-664 670 673 725-726 728 761 766-767 774 805 830 852-853 864 875 921 932 937 946 949 963 988-989 1014 1016-1017 1024 1027 1090 1097 1170 1185 1200 1215- 1216 1224 1258 1290 1309 1320 1342 1347 1355 1369 1381 1413- 1414 1431 1438 1449 1491 1512 1536 1547 1557-1560 1567 1590 1601 1636 1644 1653-1655 1662 1667 1671 1675 1680-1681 1706 1739 1760-1761 1769
fetal lung	Clontech	FLG004	103 276 334 465-466 737 843 1131 1614 1658
fetal liver-spleen	Columbia University	FLS001	3-11 13 15-21 25 30-39 41-48 50- 51 54 56-58 60-66 68-69 72 75 77-80 82-83 85 87 89 92-103 105- 110 112 116-124 126-127 130 133 135-139 141 144 147-149 152-153 157 163-165 167-172 174 176-178 180 186 188-190 193-194 196 198- 200 202-206 210-214 219 221-231 233-236 240-244 246-247 250-251 255-256 258 261-265 268-269 272 274 276-278 280-281 284-286 288 293 295 299-301 304 306-307 309 311 314 316 318 320-321 326 329- 332 342 344-345 350 352-353 356- 358 360 362 370-374 376 378-384 386-387 390 392-393 400-401 403 406 408 410-412 415 417 419 422- 437 439-442 444-445 448 452-454 456 459 461-470 472-479 481-483 487-488 490-491 493 500-501 503- 506 509-513 515-520 522-524 526- 529 531 534 536-540 542 547-549 553-554 561-562 564 567-568 571- 576 579 581 583 585-597 599-605

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			607 610-613 615-621 623-624 626 628-634 636-640 644 647-650 655- 660 665 669-670 672 674-675 678 681-682 684 690-695 697 702 708- 710 713-714 716-719 725-728 730- 731 734 736 738 740-741 743-746 748 750-751 759-766 768 772 774- 777 779 783-788 793 796 798 800- 805 808 810-812 814 818-819 821- 824 826-832 834-837 843-847 849- 867 869-876 878-883 887 889-895 897-898 902 904-914 916 919 921- 928 930-937 939 945-950 953-958 960-961 963-965 967 969 971 974- 978 980-983 986 988-990 992-993 995-997 1000-1002 1004-1008 1012 1014 1016-1019 1025-1026 1028- 1031 1033 1035-1036 1039-1044 1047 1049-1050 1053-1056 1058- 1059 1061-1064 1067-1070 1072- 1074 1076 1078 1082 1085-1087 1089-1090 1097 1099-1103 1107- 1113 1115-1119 1121-1123 1125 1127-1128 1131-1134 1136-1137 1144-1150 1153 1159-1160 1163 1170 1175 1177-1178 1188 1190- 1192 1195-1200 1202 1206 1208- 1211 1214 1216 1218 1221-1222 1225 1227 1234 1237 1241 1244 1246-1247 1251 1254 1258 1261 1266 1268 1270-1273 1277-1282 1284-1285 1287-1290 1294 1299- 1300 1306-1308 1313-1320 1324- 1325 1327 1330 1332-1333 1338 1341 1343 1345-1347 1349-1350 1353-1360 1362-1363 1365-1367 1369-1370 1372-1374 1376 1378- 1381 1383-1384 1386 1389-1391 1400 1402-1403 1405-1410 1413 1415 1417-1419 1422-1429 1431 1435-1437 1439-1442 1445-1446 1448-1449 1454 1458-1459 1466- 1470 1472 1474 1477-1478 1480 1482 1485 1491-1493 1496-1498 1501-1507 1509 1511-1512 1516- 1519 1524-1526 1529 1532 1536- 1541 1546-1547 1549-1550 1552- 1554 1562 1564 1569 1572 1574- 1575 1578 1581 1583 1587-1588 1591-1592 1594-1595 1597-1598 1600-1604 1611-1612 1614-1615 1617-1618 1620-1622 1624-1625 1627-1628 1630-1632 1634-1639 1645-1651 1653-1662 1664 1667- 1669 1671 1673-1674 1676-1688 1690 1696 1701-1703 1706-1709 1711 1713-1714 1718-1719 1722 1724-1727 1731-1733 1738 1740- 1741 1743-1744 1746 1748 1751- 1752 1754 1760-1765 1767-1773 1780 1783-1786
fetal liver-spleen	Columbia University	FLS002	3-11 13 15-21 26 29 32 35-39 42 44-45 48 50-51 54-55 57-58 61 64 68-69 73-75 78 80 82 84 87 95-98 100 103 105 107-108 110 112-113 116-119 122-125 128 130 137-138 145 147-153 155 157 159 161-163 166 168 171-172 174-175 177 181 188-189 193-194 196-198 200-203

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			206 212-215 219-221 223 225-229 231-232 240-244 246-247 250-251 258-259 262 264 268-269 272 275 277 280-281 284 286 288 290-292 295 298-299 301-304 306 308-310 318 320-321 323 325 329 331 334 342 348-349 352-353 356 359 368 371 374 376-379 381-384 386-387 392-393 397-398 400-401 403 410- 413 421 423 426-427 429-430 433- 436 438 440 443 445 448 451-452 454-455 460-463 465-467 469 471- 473 475-476 478-479 481-483 487 490-491 493-494 497 500-501 503- 505 509-513 515-517 519-520 524 526-531 534 537-542 544 547 552- 554 556 558 561-562 564-567 571- 577 583-587 590-591 593 595 597 601 604-606 608-613 616-617 619- 624 626-632 634 637-642 644 647 649-652 654-659 662-665 669-672 674-675 681-682 685 688 690 696 698 700-703 707 709-710 713 717 719-721 723-724 728 731-732 734 737-738 742-745 748 752 754 759 763-766 768 770 773-777 780 782 784 786 791 795-798 801-802 805 808 811-812 818 823-824 826-827 832 834-837 839 843 846 848-856 858-861 865 867 869 871 873-874 876 878 881-882 887 889 892 894- 898 901-902 904 906-908 913-915 919 921-924 926-932 934-935 937 939-941 943 946-947 950 953 958 961 965-967 971 973-975 977-979 981 984-985 990 992-993 995-997 999 1001 1004-1007 1009-1011 1013 1016 1020 1023 1025 1027- 1031 1033-1035 1039-1042 1044- 1045 1049 1053 1055-1056 1058- 1059 1062 1064-1065 1067-1070 1072-1074 1079 1082 1087 1089 1093 1097 1099-1103 1105-1107 1109-1114 1123 1125-1127 1132- 1134 1140 1143-1145 1148-1150 1156 1158 1160 1163 1172-1173 1177-1178 1181-1184 1190-1192 1195-1197 1199 1204 1206 1208 1211 1214 1216 1219 1227 1230 1234-1235 1237 1240-1241 1243 1245 1247 1256 1258 1260-1261 1264 1268 1270-1271 1275 1278- 1279 1284-1286 1288-1289 1299- 1301 1306 1308 1312 1314 1317- 1319 1323-1325 1327-1330 1334- 1335 1339 1343-1347 1349-1350 1354-1355 1357 1360 1362-1363 1365-1367 1369 1372 1376 1378- 1380 1386 1389-1391 1394 1400 1403 1406 1409 1416-1419 1422- 1427 1429 1435 1437-1438 1440- 1442 1446 1448-1450 1453 1460- 1461 1468 1470 1472 1474-1475 1478 1482 1486 1490-1493 1496 1498 1500-1504 1506 1508-1509 1511-1512 1516 1518-1519 1521 1524-1528 1531 1536-1538 1543 1547 1550 1554 1556 1564 1567- 1569 1580 1587-1588 1591-1592

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1597-1598 1600-1601 1611-1612 1618-1628 1630-1631 1635-1638 1641 1646-1649 1652 1654-1659 1661-1662 1664 1667-1669 1674 1676-1679 1683-1684 1686-1688 1691-1692 1699 1702 1707 1711 1713-1714 1717 1719 1722 1726- 1727 1730-1733 1738 1740 1743- 1744 1748-1752 1758 1760-1761 1763-1764 1767 1769 1772-1773 1776 1779 1783-1786
fetal liver-spleen	Columbia University	FLS003	103 300 318 321 352 372 379 381 384 392-393 403 422 424 429 434- 435 440 444 453 503 515 544 592 978 1064 1324-1325 1327 1333 1357 1369 1378 1418 1424 1622 1646 1649 1680-1681 1689-1690 1717 1743-1744 1769
fetal liver	Invitrogen	FLV001	15-16 26 34 58 61 64 70 75 78 89 98 105 112 116 120-121 123 133 151 165 176 180 194-196 198 200 204-205 210-211 220 225-226 230 235-236 239 247 259 261 267 272 277 280-281 303 310 313 317 320- 321 329 344 356 371 374 376 379- 382 395 408 412 414 419 429 434- 435 441-442 465-466 490 494 504- 506 509 522 527 534 552-553 562 567 569-570 572-574 607 631 657- 658 667 669 672 685-686 702 717 725-726 732 748 759 761 778 784 786 809 817 829 837 857 861 872- 873 875 881 889 894-895 909 911 916 954 963 967 974 977 986 988- 989 993 995 997 1000 1005-1006 1008 1014-1015 1020 1042-1043 1070 1086-1087 1089-1090 1118- 1119 1122 1144-1145 1148 1153 1157 1159 1183 1195-1196 1227 1250 1257-1258 1262 1267 1280 1285 1307 1312 1314 1317-1320 1344-1345 1349-1350 1355 1362- 1363 1403 1405 1415 1419 1425- 1426 1429 1431 1442 1448 1463- 1464 1469-1470 1489 1528 1536 1539 1549-1550 1557-1562 1577 1583 1598 1601 1611 1615 1622 1644 1649 1666 1674 1706 1721 1738 1746 1763-1765 1774 1776 1779
fetal liver	Clontech	FLV002	676 998 1719
fetal liver	Clontech	FLV004	93 133 214 301 355 374 379 555 581 601 679 837 847 859 1123 1236 1270 1313 1324-1325 1327 1355 1367 1425-1426 1536 1690 1733 1760-1761
fetal muscle	Invitrogen	FMS001	26 37-39 50-51 58 84 86 89 98 113 128 131-132 139 155 172 186 194 198 201 206 211 230-231 256 261 276 282 286 302 325 359 361 376 379 383 398 412-413 419 430 436 448 452 462-463 473 477 503 519 529 561 569-570 590-591 597 607 623 626 635 647 660 672 715 725-726 730 733 761 775-777 788 826 837 860 874 913 915 921 935 970 980 986 988-990 992 1000- 1001 1007 1014 1027 1035-1036 1045 1060 1064 1070 1083 1097

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1099-1102 1116-1117 1121 1164 1173 1198 1208 1228 1240 1258 1266 1270 1277 1298 1317-1320 1324-1325 1329 1336-1337 1369 1383-1384 1399-1400 1403 1409 1433 1505 1514 1542 1551 1554 1557-1559 1562 1589 1599 1620 1632 1644 1650 1652 1671 1675 1712 1725-1726 1743-1744 1754 1766
fetal muscle	Invitrogen	FMS002	119 221 273 402 426-427 463 547 599 736 869 1000 1033 1083 1266 1431 1440-1441 1468 1545 1599 1673 1678-1679 1687-1688 1710 1712-1714 1723 1725 1731-1733 1743-1744 1760-1761 1767
fetal skin	Invitrogen	FSK001	1 4-11 15-16 20-23 25 29 33 40 43 46 56-57 60-61 64-66 75 82 87 97-98 105 107-108 113 118-119 123 133 135-137 139 144 146 148 151-153 156 163 170 176 180 188- 189 197-198 200 202-203 210 218 222 231 246-247 261 263 265-270 277 285-286 290 293 299 301 307 311 321 325 328 330 333-335 339 341 345 351-352 355-356 358-359 362 368 370 372 376 379-382 384 388 394 404-405 408-409 411-412 419-420 424 426-427 436 441-442 445 448-449 454 462 465-466 472 476 490 493 504 506 509 515-517 519 526 531 537-540 547 549 560- 561 567 572-573 581 584 589 611- 612 615 623 630-631 635 647 649 651 657-658 660 662-665 667 669 672 676 678 681 688 701 704-705 709-710 713 717 720-721 725-726 728-729 732 748 750 753 759 764 766 770 775-777 780-781 786 788- 789 798 809 811 814 816-817 822 824-826 831 842 857 859 861 863- 864 881 894-895 908 910-911 916 918 922-923 928 932-933 935 937 946 948-949 953 960-961 966-967 970 975 977 986 990 992-993 999- 1000 1004 1007 1013 1018 1025 1027 1032 1035 1041-1043 1054 1057-1058 1060 1062-1064 1069 1072 1077 1090-1091 1097 1099- 1103 1108 1113 1119 1123 1128 1131 1134 1140 1148-1149 1152- 1153 1156 1163 1167 1178 1182 1189 1192 1195-1196 1198 1201- 1205 1208 1211-1212 1216 1219- 1220 1222 1225 1240 1243 1258 1266-1267 1274 1277 1280 1282- 1285 1299 1310 1317-1322 1324- 1325 1329-1330 1342 1344 1346 1349-1351 1354-1357 1365-1366 1369 1371 1373 1376 1378 1380 1383-1384 1387 1399-1400 1405 1410 1427 1429 1431 1433-1435 1439-1441 1448-1449 1454 1457 1468 1470 1472 1475 1480-1481 1487 1490-1491 1493 1498 1509 1512 1521 1525-1526 1529 1535- 1536 1547 1549 1557-1559 1588 1592 1595 1597-1598 1601 1603- 1604 1608 1611 1614 1618 1624-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1626 1632 1634 1636 1641 1643- 1644 1646 1654-1657 1660-1662 1665 1668 1675 1685 1687-1689 1702-1703 1709-1710 1715 1719 1724 1727 1731-1732 1737-1740 1742 1747 1749 1755 1760-1761 1765 1772 1776-1777 1779-1780 1786
fetal skin	Invitrogen	FSK002	13 286 302 307 313 321 330 335 339 341 354 370 372 385 400 402 408 414 426-427 433 436 450 454 515 544 585 598 767 810 845 939 1076 1109 1155 1317-1320 1326 1333-1335 1343 1347 1350 1369- 1371 1377-1378 1391 1397 1422 1466 1647 1656 1678-1679 1687- 1688 1693 1718 1721 1725 1731- 1732 1739 1755
fetal spleen	BioChain	FSP001	110 137 211 353 589 927 1108 1639 1771
umbilical cord	BioChain	FUC001	4-8 10 12 14 17 33-36 44-46 57 64 68-69 75 82 85 101 104 113- 114 116 119 122-124 133 137 153- 154 157 161 163 166-167 175 181- 184 186 192 197-198 200-202 212- 215 230 234 246-247 251 256 263 267 271-272 280-281 284 295 301 314 317 321 326 333-335 345 351 356 368 371-373 379-380 386 390 392 394 406 408-410 412 414 416 420 424 427 430-436 438 444-446 454 459 461 463 467 473 482-483 486 488 490 495 504 509 524 526 537-540 547 555 561 574-577 588- 591 593 606 615 620-621 632 637 645-647 650 659-660 662-664 667- 668 674-675 684 687 696 698 701 703-705 709 711 714 719-720 725- 727 732 749-750 762 765 771 775- 777 780 789-791 793 796 802-803 814-817 822 833 843 845 848 858 861 864 875 879 888 894-895 897- 900 903 906-907 911-912 925 930- 933 936 940 948 953 960 966 977 984 990 992 998 1000-1001 1005- 1007 1016 1023 1025 1037 1046- 1047 1059 1061-1063 1073 1076- 1077 1089 1094-1097 1112-1113 1115 1134 1144-1148 1151 1154 1156 1163 1171 1197 1204-1205 1208 1216 1218 1224 1234-1235 1243-1244 1246 1279 1283 1286- 1287 1298 1316 1320 1344 1346 1350 1357 1359 1371 1373 1375 1381 1398 1400 1403 1408 1414 1424 1427-1428 1431 1433 1440- 1442 1446 1454-1455 1479 1482 1484-1485 1489 1492-1493 1504- 1505 1513 1525 1527 1536 1538 1546 1565 1567 1571 1573 1575- 1576 1578-1579 1591 1595 1600- 1601 1608 1612 1615 1621 1624 1626 1636-1637 1647-1648 1651 1653 1656 1658 1661-1662 1672 1675 1682 1684 1686-1688 1690 1709-1710 1722 1727 1729 1735- 1738 1740-1741 1760-1761 1768
fetal brain	GIBCO	HFB001	4 9 11-13 17-18 22-23 25 37-39 42-47 50-51 54-55 58 60-61 65-66

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			72 75 77 80 82 85 90-91 94 100- 102 107 110 112-116 118-119 122- 123 126 128 134 136-140 147-148 153-155 157 161 165 169-172 175 181 186 188-189 197-198 204-206 208 210 215 222-223 225-226 230 235-238 240-241 247 253 256-258 260-262 267-269 276 279-281 284 286 289 298 300-302 307 310 318 321-323 325 330-331 339 341 346- 349 352 354 356-359 362 364-365 371-372 377 379-380 382 384 387 390 400 408 414-416 419 424 431 434-435 438 441-443 449 451 453- 455 457-463 470 472-473 475 477- 478 482-483 486-488 490-491 493 496 499-500 502-504 506-507 509- 512 516 519-520 522 525-526 529- 530 537-540 543-544 546-547 566- 567 569-570 572-582 585 588 590- 591 593 595 599 601 604 606-609 611-612 614-620 622-624 630 632 636 643 645-647 650-652 654 659 661 665 667-668 670-672 676 678 681 687 689 692-694 697 699 710 714 717 721 727 729-732 734 736 738 743-746 750-751 759 763 766 770 772 775-777 784 789 791 796 799 802-805 810-811 814 819-821 824 826 830 834-837 839-850 854- 856 858-860 862 864 869 871 876- 877 879 883 886-887 890-891 893- 895 898-901 905 908-910 912-916 919 922-923 925 927 930-933 935- 938 948 952-960 963-964 967 969- 972 975 978-979 981 983 986-987 990 992 995 997 999-1002 1005- 1009 1011-1013 1016 1018-1019 1023 1026 1029-1031 1033-1035 1038 1041 1047 1050 1053 1057 1059 1064 1068 1070 1072-1073 1078-1079 1081-1082 1086 1089 1094 1097 1103 1107-1109 1113- 1115 1121-1122 1127 1134-1135 1138 1140 1143 1148-1151 1153 1156-1157 1159 1167 1170 1175 1193-1194 1200 1202 1207-1209 1211 1216 1219-1220 1226-1227 1229 1232-1234 1240-1241 1243 1246 1249-1251 1253-1254 1258 1267-1268 1271 1276 1279 1282 1285-1289 1293-1294 1305 1307- 1308 1312 1316 1320 1327 1338- 1339 1341-1344 1346 1349 1355- 1357 1359 1365-1366 1369-1370 1373-1375 1379 1386 1389 1394 1398 1409 1413-1414 1416-1417 1420-1421 1425-1427 1430 1433 1437 1439 1442 1445-1452 1454- 1457 1459 1463-1464 1468 1470 1474 1477-1479 1489 1492 1494 1497-1498 1501-1503 1507 1509 1511-1513 1517 1520-1521 1524- 1526 1531-1533 1535 1537-1538 1547 1554 1556-1559 1564-1567 1571 1584 1587 1589 1594 1599- 1601 1611-1612 1614-1616 1619- 1620 1625-1628 1630-1631 1634 1637-1638 1640-1643 1645 1648

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1649 1651 1653-1655 1657-1658 1664-1665 1667 1669 1673 1678- 1679 1683-1684 1686 1693 1701 1704-1705 1709 1713-1714 1717- 1720 1724 1727-1728 1731-1733 1737-1738 1743-1744 1752 1754- 1755 1757 1760-1761 1765 1772 1779 1785
macrophage	Invitrogen	HMP001	5-8 110 204-205 503 634 678 859 878 933 988-989 1379 1448 1504
infant brain	Columbia University	IB2002	10 12-13 15-18 22-23 25 29 34 37-39 43 47 50-51 54-56 58 60-63 65-66 68-69 72-74 80 82-83 86 88-92 97 100 102-104 106-108 110 112-113 115-116 118 123 128 130 134-136 138-139 143 147-149 151- 152 154-155 163 165-167 169 172- 175 181-184 186 193-196 198 201 203-205 209-210 214-215 222 224- 226 231-232 235-236 239 246-247 252 257 260 268-269 272 276-277 279-281 286 288 291-292 295 298 300-301 304 307 310 313 321-323 330-331 333-334 339 346-347 349 352 356-357 362 371-372 377 379- 380 383-384 392 397 401 406 408 411 413-414 416 418-419 422 428 430-431 434-435 438 443 449 453- 454 461 464-466 469-470 472-473 475-476 478 482-483 487 490 492 494 497 503 507-508 510-513 516 519-520 524-526 530-534 536-540 547 550-551 561 563-564 566-567 572-576 579 581-582 584-587 590- 591 593 595-597 607-609 611-613 616-617 620 622-624 627 631 637 641 645-647 650-655 657-658 660- 665 667-675 689 691 695 697 699 703 707 713-715 717 721 728-731 733-736 739 743 745 751 755 759 763 769-770 772 778 780-781 785 788-789 793-794 799 803 808 811 814 825-826 830 834-836 840-843 845 848-850 854-855 860 862 864- 865 870 872 875-876 878 886 888 890-891 894-896 898 903-904 916- 917 919 922-925 927-928 930-932 934-936 938 941 945-946 948-950 953-954 959-962 966-969 977 979 981 986-990 992 997 999-1000 1004-1006 1014 1016 1018-1019 1024-1025 1033 1036 1047 1051- 1052 1054-1055 1057-1059 1063- 1064 1068-1070 1073 1081-1082 1085 1089 1108-1113 1118-1120 1123-1124 1130 1132-1138 1140 1149 1151 1153-1154 1163-1170 1172 1174-1175 1183-1184 1188 1190 1193-1194 1196-1197 1199 1204 1208-1209 1211 1218-1222 1226-1227 1229 1231 1234 1241 1247 1249 1251 1256 1258 1261- 1262 1269 1274 1279 1281 1283 1285 1287-1289 1294-1295 1305 1307 1313-1314 1316-1320 1329 1332 1341-1342 1345 1349 1356 1362-1363 1365-1366 1368-1370 1374 1381 1383-1384 1388 1400 1403 1406-1407 1413 1417 1420

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1423 1429-1431 1435-1436 1439- 1441 1443 1447-1449 1451-1452 1454-1455 1457 1459 1463-1465 1468 1470-1471 1475 1479 1482- 1483 1485 1493-1494 1496 1498- 1499 1502-1503 1505-1507 1509 1522-1523 1525 1528 1531-1533 1542 1546-1547 1549-1550 1554- 1555 1563 1565-1567 1569 1575 1580 1583-1586 1588 1590 1592- 1593 1595 1598 1600-1601 1608- 1610 1612 1614-1616 1619 1621 1624 1626-1627 1630-1633 1637 1639-1640 1642 1644 1647 1652 1654-1655 1658-1659 1664-1665 1672-1673 1676-1681 1685-1688 1693-1695 1701-1702 1704 1708 1717-1720 1723-1724 1726-1728 1733 1735-1741 1743-1744 1752 1755-1758 1762 1765 1771 1774 1777-1778 1786
infant brain	Columbia University	IB2003	17-18 20-23 29 34 43 50 68-69 78-80 88 100-101 107 110 112 118 123 128 133 135-137 146 148 152 159 166 169 174 194 198 203 215 223 225-226 229 235-236 247 260 276-281 286 290-292 295 300-301 310 322 324 331 334 339 346-347 349-350 352 357 371 376-377 382 384 403 408-409 414-415 453-455 472 476 478-479 490 503 507 516 520 530 534 536-540 551 563 572- 576 585 587 590-591 593 595-596 601 606 612 616-617 620 622-624 650 652-653 661 665 670-671 674- 675 678 689 715 717 727-728 730 734 759 775-777 780-781 785 796 805-807 811 824 845-846 864 869 875 882 889 894-895 898 904 917 919 921-923 932 935-936 946 950 954 962 977 979 997 999-1000 1005-1006 1009 1011 1017 1024 1033 1037 1043 1055 1057 1109 1114-1115 1120 1123 1127 1144- 1145 1149 1151-1153 1160 1167 1170 1174 1193-1194 1196 1199 1202 1206 1209 1220-1221 1226 1229 1240-1241 1251 1258 1284 1288-1289 1305 1314 1327 1333 1344 1347 1350 1356-1357 1365- 1366 1378-1379 1388 1400 1403 1421 1423 1431 1436 1440-1441 1446-1447 1457 1459 1471 1499 1503 1507 1509 1536 1546 1557- 1559 1567 1572 1587 1595 1598 1610-1612 1615 1631 1639 1644 1647 1657-1658 1673 1678-1681 1683-1684 1701-1702 1708-1709 1713-1714 1719 1757 1760-1761 1765 1771 1778
infant brain	Columbia University	IBM002	101 113 139 152 260 279 290-292 374 377 551 563 608-609 653 659 814 954 1005-1006 1029-1030 1130 1164 1209 1258 1294 1305 1320 1327 1397 1431 1498 1507 1615 1640 1694-1695 1763-1764 1767 1779
infant brain	Columbia University	IBS001	10 12 119 175 279-281 321 334 371 446 551 563 623 652 667 669

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			671-672 819 949 966 1113 1130 1151 1188 1193-1194 1196 1229 1258 1265 1271 1287 1317-1319 1324-1325 1342 1423 1440-1441 1448 1471 1482 1525 1532 1546 1562 1569 1588 1591 1610 1618 1647 1649 1658
lung, fibroblast	Strategene	LFB001	5-9 17 20-21 25 68-69 82 94 105 153 157 197-198 203 207-208 212- 213 223 262 266 283 302 321 326 333 356 370 427 430 436 446 462 472 493 498 503 516 519 527 535 537-540 542-544 562 565 567 586 599-600 607 615 630 647 662-664 692-694 712 719 745 748 775-777 794-796 810 837 843-847 849 854- 856 869 876 903 934 953 955-956 964 975-976 984 1000 1005-1007 1024-1025 1033 1039 1053 1064 1070 1072 1082 1112-1113 1134 1136-1138 1140 1195 1223 1232- 1233 1246 1279 1285 1295 1311 1320 1334-1335 1343 1427-1428 1446 1478 1482 1493 1504 1537 1552 1555 1567 1575 1582 1598 1620 1625 1632 1638 1645 1654- 1655 1662 1680-1681 1684 1686 1690 1696 1702 1711 1733 1741 1760-1761 1778 1785
lung tumor	Invitrogen	LGT002	5-10 18 20-21 29 33-36 40 43 52 54-55 61 65-66 68-70 73-75 80 85 88-89 93-94 100 103 106-108 112- 113 115-116 118-119 123-124 126 130-132 135-137 139-141 143-144 147-148 151-153 155-156 159 161 164 169 171 179-180 185 190 192 194 196-199 203-208 210 212-214 216-217 219 222 233 240-241 244 246 251-252 255-256 261-262 266. 272 276-277 279-281 284 286 288 290 295 298 301-302 309-312 317 321 329 332 341-342 344-345 348 352 358-360 363 368 370-371 376 380-381 384 389-390 398 400 409 414 423 426-427 430 432-436 443- 444 450-451 454 462 468 472-477 480-483 487-488 490-491 493 496- 498 500 503-506 509-512 515-516 519 521-523 526 530 534 541 544 547 554 557 564 566-567 572-576 585-586 588-589 595-596 601 607 611-612 615 619 621 623 626 630 632-633 644 647 649 651 655-656 660 662-665 667 669 672 683-684 696 700 706 710 713 716 718-719 722-723 728 734-739 743 750 752 763 765-766 773-778 784-785 787- 789 791 800 802-803 809-812 814 824 826 828-829 832 838-839 841- 845 849-850 852-855 857-861 864 866 874 878-880 882 887 890-891 897-898 902 904 906-907 910 916 918-920 922 924-925 927 930-932 934-935 937 947 950 953 955-956 961 963 966-967 969 971 977-979 981 984 986-987 990 992-993 995 997 999-1001 1005-1007 1009 1012-1013 1018 1020 1022-1024 1026 1029-1030 1033 1038 1041

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1045 1047-1050 1052 1054-1055 1059 1063-1064 1067-1071 1073- 1074 1078 1085 1087 1089 1095- 1097 1104 1106-1107 1109 1112 1116-1117 1119 1126 1134-1135 1139 1141-1142 1144-1145 1148 1152-1153 1156-1158 1167 1170 1172 1178 1195-1196 1198-1200 1202 1204 1208 1214 1216 1219 1222 1227 1234 1241 1247 1252 1257-1258 1265 1267-1270 1276 1278 1280-1281 1283 1285 1288- 1289 1295 1300 1305 1308 1312 1317-1321 1329 1338-1339 1341 1344-1346 1349-1351 1353-1355 1357 1365-1366 1369 1378-1379 1383-1385 1394 1397 1400 1402- 1403 1408 1417 1419 1423-1426 1431 1433-1436 1438 1444 1446- 1448 1454-1455 1460 1466 1468 1470 1474 1480-1481 1483 1486- 1488 1490-1491 1494-1496 1506 1508-1509 1511-1512 1515-1516 1519 1523-1524 1528-1529 1536- 1540 1546 1549-1550 1555 1560- 1561 1565 1567 1569 1575 1588 1591 1593-1594 1596-1598 1600- 1602 1608 1614-1616 1618 1620 1624-1625 1627-1632 1636 1639 1644-1645 1647-1649 1652-1653 1656-1662 1664 1666-1667 1670- 1671 1673-1675 1678-1679 1683 1685-1688 1690-1692 1696-1699 1705 1709 1716-1717 1722 1727 1730 1735 1739 1741 1743-1744 1748-1749 1753 1760-1762 1765 1767 1770-1771 1773 1775-1776 1778-1779 1786
lymphocytes	ATCC	LPC001	4 11-12 18 24-25 30-31 48 50-51 56-57 68-69 80 92 98 103 105 110 126 137 152-153 157 165 172 188- 189 197 203 210 217-218 222-223 225-226 229 231 247 251 256 264 272 280-281 284 300-301 321 325- 326 339 348 352 357 371 382 384 390 400 404 412 414 421 423 426- 427 430-431 445 447-448 451 454- 455 475 503 516 526-527 530 537- 540 549 556-560 563 574 577 589 602 613 615-617 621 623 628-630 636-637 647 649 657-659 690 697 717 723 755 764 775-777 780 786 789-790 793 800 802 822 838 849 866 869 876 881-883 892 898 906- 907 911 921-923 928 975 990 992 996 1001 1004-1007 1033 1050 1054 1078 1107 1135 1140-1141 1143 1148 1158 1163 1177 1199 1205 1216 1226 1231 1236 1241 1244 1250 1258 1260 1265 1269- 1271 1290-1293 1308 1312 1317 1319-1320 1339 1345-1346 1348 1350-1351 1357 1367 1369 1379 1381 1383-1384 1386-1387 1389 1394 1397 1405 1423 1425-1428 1431 1437 1446 1448 1461 1466 1470 1472 1474 1482 1492 1506 1528 1537 1546 1549 1591 1598 1600 1603-1604 1606 1627 1636

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1638 1647-1649 1651 1658-1659 1664 1676-1677 1680-1681 1687- 1688 1699 1711 1715-1716 1726 1728 1737 1740 1746 1748 1752 1756 1758 1777 1779
leukocyte	GIBCO	LUC001	3-4 10-11 13 15-18 20-21 24-25 30-31 35-36 40 43-45 48 50-51 54-58 60-63 68-69 75 79-80 82-83 85 88-91 93-96 98 100 103-104 107-108 112 116 119 123 125-128 134-140 142 147-149 151 153 155 157 162-163 167 169-172 174 177- 179 186 190 192-199 203-207 210 212-215 217-219 222-223 229 235- 236 247 251 255-258 260 262 272 274-277 280-281 285-286 297-301 307-310 313-314 316-317 321 325- 330 333-334 340-342 348-349 352 354-358 370-371 380-385 387-388 400 405 408-410 412 414-416 421- 425 430-431 434-435 437 439 441- 442 445-451 453-454 456 459 461- 464 468-472 474-479 481 483-485 487-491 496 499-501 503-504 509- 513 516-519 522 526-527 529-531 534 536-540 542 547-549 553-559 566-567 571 574-577 579 582 584- 586 589 593 595-597 601-602 604 606-607 611-613 615-621 623 627- 629 633 636-637 642 644-650 655 659-660 662-665 667 669 674-675 678 682-684 692-696 698 700 706 708 710 716-720 725-726 729-736 738-739 743-746 749 751 753 756 759 765-766 768 770-778 780 784- 786 788-790 793 796 798 800 802- 803 810-811 814 817 819 826 828- 830 832 834-836 838 843 845-860 863-864 866-871 877-879 881-892 894-896 898 902 904-914 916 919- 925 927 930-932 935-936 941-942 945 948-949 953 955-956 958 960- 962 964 967 970-971 973 975 977 985-990 992-993 995-996 999-1002 1004-1009 1011 1014 1017-1019 1022-1023 1025 1027 1029-1031 1033-1036 1038 1041 1043 1047 1050 1053-1054 1058-1059 1061- 1062 1064 1068 1070 1072 1078 1085-1086 1089-1091 1093 1097 1106-1107 1110-1113 1115-1117- 1122-1123 1125 1129 1132-1133 1135-1137 1140-1145 1152 1158 1163 1168 1170-1174 1176-1178 1180 1182-1183 1186 1195 1198- 1200 1202 1205-1206 1211 1216 1219-1221 1223-1227 1230-1236 1238-1242 1247 1252 1254 1256 1258 1261-1262 1264-1265 1269- 1270 1272-1275 1277 1280-1284 1287-1293 1299-1300 1306 1308 1312-1313 1317-1320 1322 1324- 1330 1333-1335 1339 1341 1343- 1347 1349 1353-1357 1359-1361 1365-1367 1369-1370 1373-1374 1377 1379-1381 1386-1387 1394 1400 1403 1409 1419 1423 1425- 1428 1430-1431 1433-1434 1437- 1438 1440-1442 1446-1448 1450

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1453 1458-1459 1463-1464 1468 1470-1471 1474 1477-1478 1482- 1488 1490-1493 1496-1501 1504 1506 1509 1512-1513 1516 1519 1521-1522 1524-1525 1527-1528 1531 1534 1538 1541 1545-1547 1549-1550 1553 1555-1556 1560 1565 1567 1575 1580 1589 1591 1594 1596 1598 1600-1602 1606- 1608 1611 1614 1620-1621 1624 1626-1629 1631-1632 1636 1638- 1639 1641 1644-1645 1648-1650 1653-1655 1658-1660 1662 1669- 1670 1675-1679 1684-1688 1690- 1692 1696 1700 1702 1707-1709 1711 1716-1717 1720 1723 1725- 1727 1733 1737-1738 1741 1743- 1744 1748-1749 1752 1755 1760- 1762 1763 1769 1771-1772 1781- 1784 1786
leukocyte	Clontech	LUC003	4 35-36 44-45 61 68-69 75 82 102 119 139 154 179 197 244 280-281 324 372 404 430-431 455 461 476- 477 481 503 537-540 554 575-576 581 589 608-609 621-622 624 630 632 647 662-664 669 679 698 764 773 775-777 802 848 851 856-857 879 905-907 915 949 952 990 992 1002 1113 1119 1170 1183 1216 1236-1237 1241 1275 1346 1353 1357 1359 1377 1506 1515 1534 1553 1591 1600 1613-1614 1621 1628 1670 1676-1677 1691-1692 1699 1733 1738 1772
melanoma from cell line ATCC #CRL 1424	Clontech	MEL004	25 35-36 43 80 104 126 128 150 163 166 188-189 197 210 215 220 271 277 280-281 310 317 336-338 345 351 372 380-381 383 387 412 415-416 430 445 448 454 456 467 481 490 499 503 526 528 546 548 567 575-576 588 601 613 615 647 660 665 734-735 737 759 778 787 790 800 832 845 856 859 869 878 883 887 905 914 932 934 958 976 985 990 992 999-1000 1025 1031 1038 1050 1055 1068 1074 1088 1099-1102 1107 1136-1138 1149 1156 1163 1172 1190 1195 1200 1214-1215 1217 1226-1227 1235 1238-1239 1244 1253 1278 1280 1293 1311 1320 1330 1334-1335 1345 1355 1367 1386-1387 1394 1403 1406 1414 1423 1437 1442 1465 1521 1529 1536 1539 1541 1547-1548 1582 1620 1626 1631 1638 1647 1653 1660 1667 1669- 1670 1680-1681 1696 1704 1715 1724-1725 1731-1732 1750 1760- 1761
mammary gland	Invitrogen	MMG001	5-8 10 12 14-18 20-21 24-25 29 33-39 42-43 52 55-58 60-64 68-69 71 73-74 79-80 82 89 98 100 103 106 108 112 123 128 133-137 144- 146 148 150-152 154 158-159 165- 166 170-172 174 176 178 181-185 188-190 194-198 201-206 210 217- 222 224 227-228 231 233-237 247 251 253-254 256 261-263 266-267 271 276-277 279-281 284-286 288

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			290 297 299 301 304 309-312 318 320-321 323-325 327-329 331-332 334 339 341 344-345 348 350 356 359-360 362-363 368 371 376 379- 383 388 390 393-395 397-398 405 408 412 414-415 423 430 434-437 441-444 448 451-455 462-464 474 476 479 482 485-486 488 490 494- 495 498 503 506 509-512 516-517 519-520 522 527 529 534 537-541 547 549 554 557 562 572-574 587 589-591 597 602 607 618 623 628- 629 632 634-640 644 647-648 650- 652 655 657-658 660 665 667 669- 672 674-676 679 682 688 695-696 706-707 710 713 717 720 722-730 732-734 736 738 743 747-748 750 755 759 761 766 770 780 784 786- 789 794 803 806-807 809 814 817- 822 827-829 837 842 854-858 863- 864 866 869-870 872 878 881 889 893-900 904 906-907 911 916 919 921-923 926 935-937 946 948-949 953-954 957 960-961 963 965-966 970 977-978 984-989 993-997 1000-1001 1005-1006 1008 1013- 1014 1016-1017 1023 1025 1027 1032-1033 1036 1039 1043 1045 1055 1057-1058 1063 1068-1075 1077-1078 1085 1087 1089-1091 1095-1102 1107-1108 1112-1119 1121-1123 1131-1133 1136-1137 1139-1142 1144-1145 1148-1149 1153 1159 1167 1170 1172-1173 1183-1185 1190-1192 1196-1199 1207-1208 1212 1216-1218 1222- 1223 1225 1231 1234 1240-1241 1247 1253-1254 1258-1259 1261- 1262 1270-1280 1283 1285-1286 1298 1307 1314 1316-1320 1323- 1325 1330 1334-1335 1342-1345 1349-1352 1354-1355 1359 1369- 1370 1377 1379 1381 1383-1384 1389 1405 1414 1419 1421-1423 1425-1426 1428-1429 1431 1434- 1437 1439 1448-1449 1454 1457 1460-1464 1466 1471 1480-1483 1487 1489-1491 1493 1505 1507 1512 1519 1526-1528 1532 1534 1536 1539 1542 1547 1549-1550 1554 1561-1562 1564 1567 1572 1576-1579 1581-1582 1587-1588 1592 1594 1596-1597 1601-1602 1607-1608 1610 1612-1616 1618 1621-1622 1625-1626 1631 1635- 1636 1641 1643-1644 1647 1650 1652 1654-1655 1657-1658 1660 1662 1664-1666 1669-1671 1673- 1674 1676-1677 1680-1685 1689- 1692 1701 1706 1713-1715 1719- 1720 1723-1728 1730-1732 1738 1740 1742-1744 1746-1747 1749 1751 1753 1760-1762 1765-1768 1771 1774 1776-1777 1779 1783- 1784 1786
induced neuron cells	Strategene	NTD001	29 35-36 80 116 123 156 163 181 214 230 280-281 284-285 307 321 330 340 358 371 375 377 380 382 422 424 492 497 532-533 542 546

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			549 566 586 595 612 645-647 654 734 775-778 780 792 799 821 826 856 858 875 936 953 985 990 992 1041-1043 1055 1072 1104 1193- 1194 1206 1223 1246 1253 1274 1288-1289 1291 1294 1311 1320 1349 1359 1412 1423 1485 1620 1623 1645 1684 1705 1715 1751
retinoid acid induced neuronal cells	Strategene	NTR001	5-8 78 268-269 277 383 431 506 623 677 731 999-1000 1199 1425- 1426 1547
neuronal cells	Strategene	NTU001	29 65-66 80 82 110 119 146 152 166 174 181-185 198 227-228 253 284 309 325 332 334 336-338 375 391 393 406 414-416 454 465-466 470 488 503 506 510-512 519 537- 540 572-574 597 602 607 623 647 661 700 702 716 743 771 792 858 904 948 954 977 1000 1005-1006 1025 1064 1068 1122 1148 1185 1219 1226 1234 1246 1271 1283 1295-1296 1311 1317-1320 1329- 1330 1350 1355 1365-1366 1378 1383-1384 1400 1412 1445 1505 1539 1547 1578 1647 1656 1683 1690 1738 1749 1783-1784
pituitary gland	Clontech	PIT004	311 314 379 408 419 430 454 1055 1095-1096 1272-1273 1312 1320 1378 1652 1671 1720 1725 1736 1741 1755
placenta	Clontech	PLA003	5-8 124 208 277 370 843 906-907 1280 1317-1319 1369 1609 1621 1737
prostate	Clontech	PRT001	9 46 57 71 107 147 171 177 197 201 229 231 242-243 274 280-281 307 310 317 330 358 373 382-383 400 430 434-436 451-462 469 477 489 497 500 505-506 513 521 526 531-533 547 618 649 657-658 662- 664 710 729 767 771 789 820 861 871 874 890-891 905 938 945 963- 964 988-989 1002 1025 1033 1045 1061 1095-1096 1112 1125 1142 1196 1198 1202 1232-1233 1241 1258 1272-1273 1287 1295 1313 1333 1341 1344 1349 1360 1362- 1363 1367 1437 1442 1447 1475 1478-1479 1482 1489 1513 1517 1527 1531 1536 1598-1599 1628 1636 1657 1680-1681 1687-1688 1717 1738 1743-1744
rectum	Invitrogen	REC001	17-18 29 33 62-63 71 73-74 83 86 113 126 146 153 158 167-169 195 200 206 261 309 312 341 344 368 373 388 395 408 414 420 430 441- 442 446 448 464 468 483 517 537- 540 547 567 585 589 602 623 628- 629 632 645-647 651 657-658 669 717-719 721 725-726 738 748 750 756 762-763 766 770 774 790 819 825 843 849 851 881 903 909 948- 949 960 986 996 1020 1023 1033- 1034 1064 1067 1070 1075 1086 1108-1109 1113 1130 1139 1153 1159 1172 1178 1185 1187-1189 1205 1220 1225 1240 1244 1271 1317-1320 1323 1334-1335 1350- 1351 1355 1369 1373 1375 1425-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			1426 1436 1439 1469 1474 1477 1482 1546 1587-1588 1592 1596 1610 1622 1627 1644 1658 1662 1665-1666 1669 1675-1677 1749 1786
salivary gland	Clontech	SAL001	10 55 97 103 110 140 149 152 158 198 217-218 242-243 256 301 308 312 321 333 351 354 360 410 437 448 473 487 494 496 501 535 555 569-570 572-573 590-591 624 636 651 759 762 764 768 771 788 800 809 826 848 865 879 906-907 925 933 963 1016 1020 1025 1040 1046 1055 1066 1103 1150 1172 1181 1234 1281-1282 1288-1289 1298 1315 1320 1333 1336-1337 1346 1359 1373 1379 1424 1447 1449 1474 1482 1492 1494 1498 1511 1523-1524 1537 1554 1596 1626- 1627 1636 1652-1655 1658 1665 1671-1672 1691-1692
salivary gland	Clontech	SALs03	158 326 1423 1463-1464
skin fibroblast	ATCC	SFB001	1320 1400
skin fibroblast	ATCC	SFB002	262 736 1025 1253
skin fibroblast	ATCC	SFB003	709 1119 1350 1631 1653
small intestine	Clontech	SIN001	25 142 146-147 151 155 198 203 244 260 271 280-281 286 288 298 301-302 308 312 334 340 371 398 408 412 414 416 423 426-427 430 434-435 445 452 454 478 503 516 519 521 523 543 547 549 555 559 563 569-570 585 592 604 611 626 628-629 632 650 659 681 710 714 718 750 764 780 798 829 842 857 859 866 887 892 894-895 901 904 906-907 912 919 935 997-998 1000 1007-1008 1026-1028 1044 1055 1089 1097 1116-1117 1131 1148 1169 1199 1219 1234 1247 1264 1279 1316 1320 1326 1341 1343 1349 1351 1374 1387 1398 1400 1403 1407 1423 1428 1468 1498 1501 1521 1550 1556 1585 1597 1636 1638-1639 1645 1653 1656 1662 1671 1675 1684 1691-1692 1704 1711 1717 1719 1722 1725- 1726 1729 1733-1734 1743-1744 1762 1767 1780 1785
skeletal muscle	Clontech	SKM001	18 20-21 82 84 101 118 134 148 151 153 166 225-226 258 274 277 289 329 361 412 414 424 440 452 459 470 488 503-504 537-540 647 660 673-675 715 773 780 786 830 905 922 950 963 982 990 992 1020 1047 1063 1115-1117 1121 1134 1228 1268 1284 1298 1321 1329 1336-1337 1343 1409 1413-1414 1509 1599 1624 1644 1653 1712
skeletal muscle	Clontech	SKM002	168 1683 1712
skeletal muscle	Clontech	SKMs03	235-236 1409
skeletal muscle	Clontech	SKMs04	235-236
spinal cord	Clontech	SPC001	4 9 11 17 30-31 35-36 43 46 60

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			82 85 92 94 108 110 116 139 157 167 198 204-205 210 215 229 256 259 277 280-281 300-302 304 315 317 372 379 387 392 419 426-427 430 433 448 467 473 487 489 506 509 513 519 524 526 537-540 543 547 549 551 559 567 569-570 593 607 616-617 623 625 637 649-650 652 657-658 670-671 673 679 681- 682 709 711 715 719 728-729 734 749-750 753 775-777 781 789 791 809 820 832 834-836 847-849 854- 855 858 861 864 871-872 875 884 898 906-908 917 919 924 934 942 944 970 985 990 992-993 998 1013 1039 1053 1059 1065 1072 1075 1077 1082 1085 1097 1103 1109 1116-1117 1128 1134 1151 1170 1174 1192-1194 1215 1225 1241 1243 1283 1294 1307 1312 1320 1323 1327 1330 1350 1353-1354 1356 1359 1368 1375 1400 1406- 1407 1423 1429 1437 1443 1448 1454 1470 1482 1492 1501 1508 1511 1529 1538 1548-1549 1565 1571 1578 1598 1600 1614 1625 1627 1630 1639 1646 1651-1652 1670 1686 1696 1740 1751 1755 1771
adult spleen	Clontech	SPLc01	117 312 326 348 424 426-427 431 845 866 1320 1330 1333 1344 1355-1357 1371 1387 1397 1446 1538 1579 1669 1686 1739 1767
stomach	Clontech	ST0001	10 15-16 61 68-69 100 117 149 197 201 227-228 231 249 273 280- 281 287 291-292 302 312 358 362 426-427 430 446 462 475 479 535 597 620 630 651 662-664 722 739 780 782 785 846 919 960 964 966- 967 976 1008 1012 1032 1042 1063 1071 1135 1170 1208 1234-1235 1259 1277 1280-1281 1322 1349 1359 1369 1449 1468 1474 1478 1487 1493 1498 1557-1559 1622 1634 1651 1653 1729
thalamus	Clontech	THA002	9 11 25 85 87 112 137 146 180 190 198 206 210 212-213 235-236 239 261 268-269 279 290 301 325 333-334 341 351 356 364-365 379 388 393 396 419-420 441-442 458 477 483 508 525 531 549 567 606 608-609 647 681 715 725-727 736 774 782 784 794 827 883 890-891 899-900 961 997 999-1001 1004 1034 1055 1097 1129 1144-1145 1150-1151 1157 1172-1173 1177 1193-1194 1208 1220 1249 1280 1305 1345 1355 1369 1434-1435 1440-1441 1454 1496 1546 1549 1562 1572 1578 1590 1594 1613- 1614 1640 1651-1652 1671 1687- 1688 1703 1743-1744 1746-1747 1753
thymus	Clontech	THM001	44-45 54 57-58 62-64 79 104 123 126 134 153 193 212-213 218 242- 243 258 274 277 279 297 301 307 327 330 333 342 351 358 371 410 430 445 465-466 468 471 483 487 493 503 506 509 517 526 535 537-

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			540 546 548 554 567 584 586 590- 591 604 612 621 638-640 645-647 649 656 660 665 670 698 710 720 728 735 739 746 759 762 766-767 775-777 780 784-785 800 802 809 824 826 828 845 851 858-859 864 866 870-871 878 884 887 892 899- 900 927 930-931 967 983 986 990 992 999 1014 1029-1030 1033 1059 1066 1073 1103 1107 1113 1116- 1117 1119 1140-1142 1158 1163 1172 1177 1195 1206 1209 1213 1216 1218-1219 1221-1222 1227 1271 1277 1282 1320 1329 1349 1367 1369 1383-1384 1417 1419 1423 1425-1427 1448 1477 1488 1493 1536 1554 1620 1644 1646 1649 1654-1655 1661-1662 1669- 1670 1674 1676-1677 1685-1688 1707 1711 1731-1732 1737
thymus	Clontech	THMc02	5-9 15-21 25 33 35-36 43-45 48 50-51 54-55 60 75 83 87 89 93 98-100 102 105 112 117 135-137 141 143 146 157 167 169 192 196 211 217-219 222 224 229 233 235- 236 240-241 244 251-252 256 261- 262 268-269 286 288 290 295 297 301-302 309-310 315-317 321 324 327 334 342 350 352-353 360 370- 373 382 384 400 403 410 414-416 424 430-431 436 445 454-456 461 464-467 470 472 474-476 483 488 497 500 504 506 513 516 519-520 524 526 530-531 534 537-540 549 554-555 565-566 569-570 572-573 575-577 586-587 595 603-604 606 612 630-632 634 636 647 650 657- 660 666-667 669 673-675 678 698 700 703 708 720 725-726 731 738- 739 743-744 750-753 757 759 763- 765 767 772-779 787 789-790 798 800 810 823 829 834-836 841 848 854-856 859 861 864 870-871 881 890-891 898 908-909 913 928 933 941 949 958 961 963 967 969 975 981 986 988-990 992 999 1007- 1008 1014 1016 1039 1041 1073- 1074 1079 1089 1097 1109 1114- 1117 1122 1131 1140-1141 1144- 1145 1163 1172 1175-1177 1186 1196 1198 1206 1211 1216 1220 1223 1227 1234-1243 1261-1262 1267 1271 1280-1281 1284 1290 1308 1317-1320 1322 1324-1325 1327 1330 1334-1335 1339 1346 1350-1351 1355 1357 1360 1370 1374 1377-1379 1386 1389-1390 1392 1397 1400 1402 1406-1407 1417 1423 1425-1427 1440-1441 1466 1474 1477 1483 1493 1498 1504 1506 1525 1536 1545 1549 1566 1594 1598-1600 1608 1611 1614 1621 1623 1625 1632 1639 1641 1644 1647 1649 1653-1656 1658 1662-1663 1671 1673 1678- 1681 1686-1688 1693 1705 1707 1711 1717-1718 1726-1727 1731- 1733 1737-1738 1743-1745 1758- 1761 1771-1772 1779 1786

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
thyroid gland	Clontech	THR001	4 9-10 20-21 37-39 48 50-51 54-57 60-61 65-66 71 83 94-96 98-100 102 104 110 112 115-117 119 123 127 133 136-137 140 149 152-153 155-158 163-164 168-169 171 186 190-192 197 201-203 219-220 229 233-237 246-247 253 256 258 262 265-266 268-269 277 280-281 284-286 288-289 298-299 302 309-311 317 321 326 332 335 341-342 344 348 350 354 358-359 363 368 371-373 382-383 385 394 398 400-401 411 414-415 421 424 430-431 433-436 443-446 450-452 454-455 458 472-474 476-478 482 484-485 487-488 490-494 496-497 500-501 503-504 506 509-513 516-517 519 524 526-527 529 535-540 547 549 562 564 569-570 575-576 588 594-595 601-602 604 606 610 612 615-617 619-623 628-630 634-635 642 647 649-651 660 662-665 668 670 681 690-694 696 698 700 709 721 727-729 732 734 738 740-741 743 745 750 759 761 763 765 770 773 780 785 795-796 798 802 804 823-824 826 828 833 838 841-845 847 849 857-860 867 874-875 878 880-881 887-888 890-892 894-895 898 908 910-911 913-914 922-923 926-927 929 932-934 937 939 941-942 948 953 957 961 963-964 966 978-979 981-982 987 990 992 1001 1004-1006 1010 1014 1020 1024 1033 1038-1039 1044 1047 1050 1052-1054 1056 1058 1068 1070-1071 1077-1079 1088 1094-1097 1105-1106 1112-1113 1116-1117 1124 1126 1128-1129 1131 1134 1136-1137 1142-1143 1146-1147 1149-1150 1156 1161-1164 1167 1170-1173 1177-1181 1190 1192 1197 1200 1204 1208-1209 1214 1217 1219 1222 1230 1232-1233 1235 1241 1245 1247 1254 1257-1258 1260 1262 1271-1273 1283 1286-1289 1299 1306 1314 1320 1330-1332 1334-1335 1342 1345 1349 1365-1367 1370-1372 1374 1381 1394 1407 1419 1428 1436-1437 1440-1441 1443 1446-1449 1454 1459 1461-1462 1468 1470-1471 1475 1477 1479 1482 1491 1497-1498 1504-1505 1507 1513 1522 1524-1526 1528 1531 1534 1536-1537 1548 1550 1553 1555-1559 1562 1567 1578 1590-1591 1597 1599-1601 1612 1614 1616 1619-1620 1622 1624-1626 1628 1631-1632 1634 1636 1639 1644-1645 1648 1651 1653-1656 1658 1660 1662-1663 1667 1669 1671 1675 1678-1681 1683-1686 1689 1691-1692 1703 1709-1711 1717 1724-1726 1729 1734 1737-1738 1740 1743-1744 1749 1753 1759-1761 1770 1777 1786
trachea	Clontech	TRC001	9 29-31 46 48 87 104 107 110 135 158 222 262 266 286 301 318 331

Tissue Origin	RNA Source	Hyseq Library Name	SEQ ID NOS:
			352 372 377 384 414 424 424 445-446 454 472 474 491 496 560 579 588 593 597 607 612 626 681 702 719 810 859 866 878 894-895 912 916 922 932 935 1046 1075 1080 1099- 1102 1113 1208 1215 1232-1233 1237 1281 1312 1385 1387 1405 1414 1424 1430 1437 1447 1505 1569 1579 1586 1600 1641 1653 1667 1671 1676-1677 1683 1691- 1692 1711 1717 1726 1772
uterus	Clontech	UTR001	17 19 25 41 46 57-58 61 89 104 108 139 152 174 198 200-201 206 263-265 274 290 387 408 420 438 446 448 452 473 491 493 499 503 506 513 519 522 526 530 542-543 560 601 610 632 659 665 720 751 773 780 833 845 857 872 877 912 929 934 937 996 1009-1011 1018 1050 1075 1107 1124 1170 1219 1258 1279 1287 1310 1320 1323 1343-1344 1375 1437 1451-1452 1478 1481 1498 1519 1521 1536 1552 1579 1597 1602 1606 1620 1626-1627 1649 1652 1661 1670 1719 1722-1723

TRADOC:1416191.1(%CQN011.DOC)

TABLE 6

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number corresponding SEQ ID NO: in priority application	SEQ ID NO: in U.S.S.N. 09/488,725
1	1787	3573	5359	784CIP2_1	1103
2	1788	3574	5360	784CIP2_2	2673
3	1789	3575	5361	784CIP2_3	4117
4	1790	3576	5362	784CIP2_4	5556
5	1791	3577	5363	784CIP2_5	5562
6	1792	3578	5364	784CIP2_6	5562
7	1793	3579	5365	784CIP2_7	5562
8	1794	3580	5366	784CIP2_8	5562
9	1795	3581	5367	784CIP2_9	5563
10	1796	3582	5368	784CIP2_10	5564
11	1797	3583	5369	784CIP2_11	5565
12	1798	3584	5370	784CIP2_12	5689
13	1799	3585	5371	784CIP2_13	5729
14	1800	3586	5372	784CIP2_14	5745
15	1801	3587	5373	784CIP2_15	5777
16	1802	3588	5374	784CIP2_16	5777
17	1803	3589	5375	784CIP2_17	5789
18	1804	3590	5376	784CIP2_18	5792
19	1805	3591	5377	784CIP2_19	5804
20	1806	3592	5378	784CIP2_20	5805
21	1807	3593	5379	784CIP2_21	5805
22	1808	3594	5380	784CIP2_22	5844
23	1809	3595	5381	784CIP2_23	5844
24	1810	3596	5382	784CIP2_24	5850
25	1811	3597	5383	784CIP2_25	5867
26	1812	3598	5384	784CIP2_26	5973
27	1813	3599	5385	784CIP2_27	5995
28	1814	3600	5386	784CIP2_28	5995
29	1815	3601	5387	784CIP2_29	6005
30	1816	3602	5388	784CIP2_30	6007
31	1817	3603	5389	784CIP2_31	6007
32	1818	3604	5390	784CIP2_32	6009
33	1819	3605	5391	784CIP2_33	6012
34	1820	3606	5392	784CIP2_34	6015
35	1821	3607	5393	784CIP2_35	6016
36	1822	3608	5394	784CIP2_36	6016
37	1823	3609	5395	784CIP2_37	6018
38	1824	3610	5396	784CIP2_38	6018
39	1825	3611	5397	784CIP2_39	6018
40	1826	3612	5398	784CIP2_40	6023
41	1827	3613	5399	784CIP2_41	6070
42	1828	3614	5400	784CIP2_42	6081
43	1829	3615	5401	784CIP2_43	6089
44	1830	3616	5402	784CIP2_44	6118
45	1831	3617	5403	784CIP2_45	6118
46	1832	3618	5404	784CIP2_46	6130
47	1833	3619	5405	784CIP2_47	6177
48	1834	3620	5406	784CIP2_48	6189
49	1835	3621	5407	784CIP2_49	6191
50	1836	3622	5408	784CIP2_50	6204
51	1837	3623	5409	784CIP2_51	6204
52	1838	3624	5410	784CIP2_52	6284
53	1839	3625	5411	784CIP2_53	6367
54	1840	3626	5412	784CIP2_54	6436
55	1841	3627	5413	784CIP2_55	6442
56	1842	3628	5414	784CIP2_56	6445
57	1843	3629	5415	784CIP2_57	6457
58	1844	3630	5416	784CIP2_58	6458
59	1845	3631	5417	784CIP2_59	6458

SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number corresponding SEQ ID NO: in priority application	SEQ ID NO: in U.S.S.N. 09/488,725
1114	2900	4686	6472	784CIP2B_803	8038
1115	2901	4687	6473	784CIP2B_804	8042
1116	2902	4688	6474	784CIP2B_805	8045
1117	2903	4689	6475	784CIP2B_806	8045
1118	2904	4690	6476	784CIP2B_807	8046
1119	2905	4691	6477	784CIP2B_808	8047
1120	2906	4692	6478	784CIP2B_809	8051
1121	2907	4693	6479	784CIP2B_810	8059
1122	2908	4694	6480	784CIP2B_811	8064
1123	2909	4695	6481	784CIP2B_812	8069
1124	2910	4696	6482	784CIP2B_813	8074
1125	2911	4697	6483	784CIP2B_814	8077
1126	2912	4698	6484	784CIP2B_815	8078
1127	2913	4699	6485	784CIP2B_816	8079
1128	2914	4700	6486	784CIP2B_817	8084
1129	2915	4701	6487	784CIP2B_818	8088
1130	2916	4702	6488	784CIP2B_819	8090
1131	2917	4703	6489	784CIP2B_820	8091
1132	2918	4704	6490	784CIP2B_821	8099
1133	2919	4705	6491	784CIP2B_822	8099
1134	2920	4706	6492	784CIP2B_823	8100
1135	2921	4707	6493	784CIP2B_824	8102
1136	2922	4708	6494	784CIP2B_825	8103
1137	2923	4709	6495	784CIP2B_826	8103
1138	2924	4710	6496	784CIP2B_827	8104
1139	2925	4711	6497	784CIP2B_828	8108
1140	2926	4712	6498	784CIP2B_829	8110
1141	2927	4713	6499	784CIP2B_830	8116
1142	2928	4714	6500	784CIP2B_831	8117
1143	2929	4715	6501	784CIP2B_832	8123
1144	2930	4716	6502	784CIP2B_833	8130
1145	2931	4717	6503	784CIP2B_834	8130
1146	2932	4718	6504	784CIP2B_835	8143
1147	2933	4719	6505	784CIP2B_836	8143
1148	2934	4720	6506	784CIP2B_837	8154
1149	2935	4721	6507	784CIP2B_838	8155
1150	2936	4722	6508	784CIP2B_839	8162
1151	2937	4723	6509	784CIP2B_840	8163
1152	2938	4724	6510	784CIP2B_841	8172
1153	2939	4725	6511	784CIP2B_842	8173
1154	2940	4726	6512	784CIP2B_843	8179
1155	2941	4727	6513	784CIP2B_844	8182
1156	2942	4728	6514	784CIP2B_845	8183
1157	2943	4729	6515	784CIP2B_846	8184
1158	2944	4730	6516	784CIP2B_847	8185
1159	2945	4731	6517	784CIP2B_848	8187
1160	2946	4732	6518	784CIP2B_849	8188
1161	2947	4733	6519	784CIP2B_850	8190
1162	2948	4734	6520	784CIP2B_851	8190
1163	2949	4735	6521	784CIP2B_852	8192
1164	2950	4736	6522	784CIP2B_853	8193
1165	2951	4737	6523	784CIP2B_854	8197
1166	2952	4738	6524	784CIP2B_855	8197
1167	2953	4739	6525	784CIP2B_856	8199
1168	2954	4740	6526	784CIP2B_857	8202
1169	2955	4741	6527	784CIP2B_858	8203
1170	2956	4742	6528	784CIP2B_859	8208
1171	2957	4743	6529	784CIP2B_860	8209
1172	2958	4744	6530	784CIP2B_861	8211
1173	2959	4745	6531	784CIP2B_862	8214
1174	2960	4746	6532	784CIP2B_863	8217
1175	2961	4747	6533	784CIP2B_864	8223

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, and complementary sequences thereof.
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

- (a) a polypeptide encoded by any one of the polynucleotides of claim 1; and
- (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO:1-1786 and 3573-5358.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

- a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and
- b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

- a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;
- b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and
- c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

- a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and
- b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

- a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and
- b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

- a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and
- b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

- a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO:1-1786 and 3573-5358, under conditions sufficient to express the polypeptide in said cell; and
- b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO:1787 -3572 and 5359-7144, the mature protein portion thereof, or the active domain thereof.
21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.
22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO:1-1786 and 3573-5358.
23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.
24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.
25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.
26. The collection of claim 22, wherein the collection is provided in a computer-readable format.
27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.
28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

SEQUENCE LISTING

<110> Hyseq Inc
<120> Novel Nucleic Acid and Polypeptides
<130> 784FLPCT
<140> To be assigned
<141>
<150> US/09/488,725
<151> 2000-01-21
<150> US09/552,317
<151> 2000-04-25
<150> US09/598,042
<151> 2000-06-20
<150> US09/620,312
<151> 2000-07-19
<150> US09/653,450
<151> 2000-08-31
<150> US09/662,191
<151> 2000-09-14
<150> US09/693,036
<151> 2000-10-19
<150> US09/727,344
<151> 2000-11-29

<160> 7144

<170> pt_FL_genes_b Versions 1.0

<210> 1
<211> 1630
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (261)..(1052)

<400> 1
cccgacgtcg catgctcccg gccgccatgg ccgcgggatt ttcccggtc gacgatttcg 60
ttttgctcac aggctggagt gcagtgggc aatctcagct tactgcaacc tccgcccagtg 120
aggcccttca ttagacacta atttgaagct atggcattcc cccactatga gcccacctgt 180
cagccaggct atctgccttg atccttagatg aagtggccat tctgcctgcc cctcagaacc 240

aat cat ctc ctc act gat att ctt gct gat gtg aag aca aaa aga aaa Asn His Leu Leu Thr Asp Ile Leu Ala Asp Val Lys Thr Lys Arg Lys 545 550 555	2100
act ttg gca aat aat aaa tta cat caa atg gaa aga gaa ttc tat gta Thr Leu Ala Asn Asn Lys Leu His Gln Met Glu Arg Glu Phe Tyr Val 560 565 570	2148
tat ttt tta aaa gat gaa gat tat ctg aaa gat att gtg gag aat tta Tyr Phe Leu Lys Asp Glu Asp Tyr Leu Lys Asp Ile Val Glu Asn Leu 575 580 585 590	2196
gaa act caa tca aag att aag gct gtt agt ctt gaa gat tga aaattac Glu Thr Gln Ser Lys Ile Lys Ala Val Ser Leu Glu Asp * 595 600	2245
tgaaaaactga atctttatta cgtgtcctct tttatttattt agaagactgt gtataataaa	2305
cactactaaa tttttaaaaaa aaaaaaaaa	2332

<210> 1172
 <211> 4494
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (207) .. (3002)

<400> 1172	
attcccggtt cgacgatttc gtactgcgga ttcaggaggg aatgaagccg gggggcccaag ctccctgtgag ccgcagctgt gtctctccat ctccctgggtc ctatgagctg cggctgtttc tctccatctc ctgggtcctg tgaggcgcag cccatcatgg ggccattgtt caggtgcacc ctttgtccc gggctcatcc accatc atg atc cat gac ttg tgc ctc gtc ttc Met Ile His Asp Leu Cys Leu Val Phe 1 5	60
ccg gcc cca gcc aag gct gtc gtt tac gtg tcc gac att cag gag ctg Pro Ala Pro Ala Lys Ala Val Val Tyr Val Ser Asp Ile Gln Glu Leu 10 15 20 25	120
tac atc cgt gtg gtt gac aag gtg gag att ggg aag aca gtg aag gca Tyr Ile Arg Val Val Asp Lys Val Glu Ile Gly Lys Thr Val Lys Ala 30 35 40	180
tac gtc cgc gtg ctg gac ttg cac aag aag ccc ttc ctt gcc aaa tac Tyr Val Arg Val Leu Asp Leu His Lys Lys Pro Phe Leu Ala Lys Tyr 45 50 55	233
tcc ccc ttt atg gac ctg aag ctc cga gca gcc tcc ccc atc att aca Phe Pro Phe Met Asp Leu Lys Leu Arg Ala Ala Ser Pro Ile Ile Thr 60 65 70	281
ttg gtg gcc ctt gat gaa gcc ctt gac aac tac acc atc aca ttc ctc Leu Val Ala Leu Asp Glu Ala Leu Asp Asn Tyr Thr Ile Thr Phe Leu 75 80 85	329
	377
	425
	473

atc cgc ggt gtg gcc atc ggc cag acc agt cta act gca agt gtg acc Ile Arg Gly Val Ala Ile Gly Gln Thr Ser Leu Thr Ala Ser Val Thr 90 95 100 105	521
aat aaa gct gga cag aga atc aac tca gcc cca caa cag att gaa gtc Asn Lys Ala Gly Gln Arg Ile Asn Ser Ala Pro Gln Gln Ile Glu Val 110 115 120	569
ttt ccc ccg ttc agg ctg atg ccc agg aag gtg aca ctg ctt atc ggg Phe Pro Pro Phe Arg Leu Met Pro Arg Lys Val Thr Leu Leu Ile Gly 125 130 135	617
gcc acg atg cag gtc acc tcc gag ggc ggc ccc cag cct cag tcc aac Ala Thr Met Gln Val Thr Ser Glu Gly Gly Pro Gln Pro Gln Ser Asn 140 145 150	665
atc ctt ttc tcc atc agc aat gag agc gtt gcg ctg gtg agc gct gct Ile Leu Phe Ser Ile Ser Asn Glu Ser Val Ala Leu Val Ser Ala Ala 155 160 165	713
ggg ctg gta cag ggc ctc gcc atc ggg aac ggc act gtg tct ggg ctc Gly Leu Val Gln Gly Leu Ala Ile Gly Asn Gly Thr Val Ser Gly Leu 170 175 180 185	761
gtg cag gca gtg gat gca gag acc ggc aag gtg gtc atc atc tct cag Val Gln Ala Val Asp Ala Glu Thr Gly Lys Val Val Ile Ile Ser Gln 190 195 200	809
gac ctc gtg cag gtg gag gtg ctg ctg 'cta agg gcc gtg agg atc cgc Asp Leu Val Gln Val Glu Val Leu Leu Arg Ala Val Arg Ile Arg 205 210 215	857
gcc ccc atc atg cgg atg agg acg ggc acc cag atg ccc atc tat gtc Ala Pro Ile Met Arg Met Arg Thr Gly Thr Gln Met Pro Ile Tyr Val 220 225 230	905
acc ggc atc acc aac cac cag aac cct ttc tcc ttt ggc aat gcc gtg Thr Gly Ile Thr Asn His Gln Asn Pro Phe Ser Phe Gly Asn Ala Val 235 240 245	953
cca ggc ctg acc ttc cac tgg tct gtc acc aag cgg gac gtc ctg gac Pro Gly Leu Thr Phe His Trp Ser Val Thr Lys Arg Asp Val Leu Asp 250 255 260 265	1001
ctc cga ggg cgg cac cac gag gcg tcg atc cga ctc ccg tca cag tac Leu Arg Gly Arg His His Glu Ala Ser Ile Arg Leu Pro Ser Gln Tyr 270 275 280	1049
aac ttt gcc atg aac gtg ctc ggc cgg gta aaa ggc cgg acc ggg ctg Asn Phe Ala Met Asn Val Leu Gly Arg Val Lys Gly Arg Thr Gly Leu 285 290 295	1097
agg gcg gtg gtc aag gct gtg gac ccc aca tcg ggg cag ctg tat ggc Arg Ala Val Val Lys Ala Val Asp Pro Thr Ser Gly Gln Leu Tyr Gly 300 305 310	1145
ctg gcc aga gaa ctc tcg gat gag atc caa gtc cag gtg ttt gag aag Leu Ala Arg Glu Leu Ser Asp Glu Ile Gln Val Gln Val Phe Glu Lys 315 320 325	1193
ctg cag ctg ctc aac cct gaa ata gaa gca gaa caa ata tta atg tcg Leu Gln Leu Leu Asn Pro Glu Ile Glu Ala Glu Gln Ile Leu Met Ser 330 335 340 345	1241

ccc aac tca tat ata aag ctg cag aca aac agg gat ggt gca gcc tct Pro Asn Ser Tyr Ile Lys Leu Gln Thr Asn Arg Asp Gly Ala Ala Ser 350 355 360	1289
ctg agc tac cgc gtc ctg gat gga ccc gaa aag gtt cca gtt gtg cat Leu Ser Tyr Arg Val Leu Asp Gly Pro Glu Lys Val Pro Val Val His 365 370 375	1337
gtt gat gag aaa ggc ttt cta gca tca ggg tct atg atc ggg aca tcc Val Asp Glu Lys Gly Phe Leu Ala Ser Gly Ser Met Ile Gly Thr Ser 380 385 390	1385
acc atc gaa gtg att gca caa gag ccc ttt ggg gcc aac caa acc atc Thr Ile Glu Val Ile Ala Gln Glu Pro Phe Gly Ala Asn Gln Thr Ile 395 400 405	1433
att gtt gct gta aag gta tcc cct gtt tcc tac ctg agg gtt tcc atg Ile Val Ala Val Lys Val Ser Pro Val Ser Tyr Leu Arg Val Ser Met 410 415 420 425	1481
agc cct gtc ctg cac acc cag aac aag gag gcc ctg gtg gcc gtg cct Ser Pro Val Leu His Thr Gln Asn Lys Glu Ala Leu Val Ala Val Pro 430 435 440	1529
ttg gga atg acc gtg acc ttc act gtc cac ttc cac gac aac tct gga Leu Gly Met Thr Val Thr Phe Thr Val His Phe His Asp Asn Ser Gly 445 450 455	1577
gat gtc ttc cat gct cac agt tcg gtc ctc aac ttt gcc act aac aga Asp Val Phe His Ala His Ser Ser Val Leu Asn Phe Ala Thr Asn Arg 460 465 470	1625
gac gac ttt gtg cag atc ggg aag ggc ccc acc aac aac acc tgt gtt Asp Asp Phe Val Gln Ile Gly Lys Gly Pro Thr Asn Asn Thr Cys Val 475 480 485	1673
gtc cgc aca gtc agc gtg ggc ctg aca ctg ctc cgt gtg tgg gac gca Val Arg Thr Val Ser Val Gly Leu Thr Leu Leu Arg Val Trp Asp Ala 490 495 500 505	1721
gag cac ccg ggc ctc tcg gac ttc atg ccc ctg cct gtc cta cag gcc Glu His Pro Gly Leu Ser Asp Phe Met Pro Leu Pro Val Leu Gln Ala 510 515 520	1769
atc tcc cca gag ctg tct ggg gcc atg gtg gtg ggg gac gtg ctc tgt Ile Ser Pro Glu Leu Ser Gly Ala Met Val Val Gly Asp Val Leu Cys 525 530 535	1817
ctg gcc act gtt ctg acc agc ctg gaa ggc ctc tca gga acc tgg agc Leu Ala Thr Val Leu Thr Ser Leu Glu Gly Leu Ser Gly Thr Trp Ser 540 545 550	1865
tcc tcg gcc aac agc atc ctc cac atc gac ccc aag acg ggt gtg gct Ser Ser Ala Asn Ser Ile Leu His Ile Asp Pro Lys Thr Gly Val Ala 555 560 565	1913
gtg gcc cgg gcc gtg gga tcc gtg acg gtt tac tat gag gtc gct ggg Val Ala Arg Ala Val Gly Ser Val Thr Val Tyr Tyr Glu Val Ala Gly 570 575 580 585	1961
cac ctg agg acc tac aag gag gtg gtg gtc agc gtc cct cag agg atc His Leu Arg Thr Tyr Lys Glu Val Val Ser Val Pro Gln Arg Ile 590 595 600	2009

atg gcc cgt cac ctc cac ccc atc cag acc agc ttc cag gag gct aca Met Ala Arg His Leu His Pro Ile Gln Thr Ser Phe Gln Glu Ala Thr 605 610 615	2057
gcc tcc aaa gtg att gtt gcc gtg gga gac aga agc tct aac ctg aga Ala Ser Lys Val Ile Val Ala Val Gly Asp Arg Ser Ser Asn Leu Arg 620 625 630	2105
ggc gag tgc acc ccc acc cag agg gaa gtc atc cag gcc ttg cac cca Gly Glu Cys Thr Pro Thr Gln Arg Glu Val Ile Gln Ala Leu His Pro 635 640 645	2153
gag acc ctc atc agc tgc cag tcc cag ttc aag ccg gcc gtc ttt gat Glu Thr Leu Ile Ser Cys Gln Ser Gln Phe Lys Pro Ala Val Phe Asp 650 655 660 665	2201
ttc cca tct caa gat gtg ttc acc gtg gag cca cag ttt gac act gct Phe Pro Ser Gln Asp Val Phe Thr Val Glu Pro Gln Phe Asp Thr Ala 670 675 680	2249
ctc ggc cag tac ttc tgc tca atc aca atg cac agg ctg acg gac aag Leu Gly Gln Tyr Phe Cys Ser Ile Thr Met His Arg Leu Thr Asp Lys 685 690 695	2297
cag cgg aag cac ctg agc atg aag aag aca gct ctg gtg gtc agt gcc Gln Arg Lys His Leu Ser Met Lys Lys Thr Ala Leu Val Val Ser Ala 700 705 710	2345
tcc ctc tcc agc agc cac ttc tcc aca gag cag gtg ggg gcc gag gtg Ser Leu Ser Ser His Phe Ser Thr Glu Gln Val Gly Ala Glu Val 715 720 725	2393
ccc ttc agc cca ggt ctc ttc gcc gac cag gct gaa atc ctt ttg agc Pro Phe Ser Pro Gly Leu Phe Ala Asp Gln Ala Glu Ile Leu Leu Ser 730 735 740 745	2441
aac cac tac acc agt tcc gag atc agg gtc ttt ggt gcc ccg gag gtt Asn His Tyr Thr Ser Ser Glu Ile Arg Val Phe Gly Ala Pro Glu Val 750 755 760	2489
ctg gag aac ttg gag gtg aaa tcc ggg tcc ccg gcc gtg ctg gca ttc Leu Glu Asn Leu Glu Val Lys Ser Gly Ser Pro Ala Val Leu Ala Phe 765 770 775	2537
gca aag gag aag tct ttt ggg tgg ccc agc ttc atc aca tac acg gtc Ala Lys Glu Lys Ser Phe Gly Trp Pro Ser Phe Ile Thr Tyr Thr Val 780 785 790	2585
ggc gtc tcg gac ccc gcg gct ggc agc caa ggg cct ctg tcc act acc Gly Val Ser Asp Pro Ala Ala Gly Ser Gln Gly Pro Leu Ser Thr Thr 795 800 805	2633
ctg acc ttc tcc agc ccc gtg acc aac caa gcc att gcc atc cca gtg Leu Thr Phe Ser Ser Pro Val Thr Asn Gln Ala Ile Ala Ile Pro Val 810 815 820 825	2681
aca gtg gct ttt gtg atg gat cgc cgt ggg ccc ggt cct tat gga gcc Thr Val Ala Phe Val Met Asp Arg Arg Gly Pro Gly Pro Tyr Gly Ala 830 835 840	2729
agc ctc ttc cag cac ttc ctg gat tcc tac cag gtc atg ttc ttc acg Ser Leu Phe Gln His Phe Leu Asp Ser Tyr Gln Val Met Phe Phe Thr 845 850 855	2777

ctc ttc gcc ctg ttg gct ggg aca gca gtc atg atc ata gcc tac cac	2825
Leu Phe Ala Leu Leu Ala Gly Thr Ala Val Met Ile Ile Ala Tyr His	
860 865. 870	
act gtc tgc acg ccc cgg gat ctt gct gtg cct gca gcc ctc acg cct	2873
Thr Val Cys Thr Pro Arg Asp Leu Ala Val Pro Ala Ala Leu Thr Pro	
875 880 885	
cga gcc agc cct gga cac agc ccc cac tat ttc gct gcc tca tca ccc	2921
Arg Ala Ser Pro Gly His Ser Pro His Tyr Phe Ala Ala Ser Ser Pro	
890 895 900 905	
aca tct ccc aat gca ttg cct cct gct cgc aaa gcc agc cct ccc tca	2969
Thr Ser Pro Asn Ala Leu Pro Pro Ala Arg Lys Ala Ser Pro Pro Ser	
910 915 920	
ggg ctg tgg agc cca gcc tat gcc tcc cac tag gccgcgtg aagggtccccg	3020
Gly Leu Trp Ser Pro Ala Tyr Ala Ser His *	
925 930	
gaggatgggt ctcagccgag cctcgtgcac ccccaagatg gaacatccct gctgcattca	3080
cactggaaaca agccccctcca gatgagtgcc ccggcccccag gccagcttca ctgcccgtctc	3140
ttcacacaga gctgttagttt cggctctgcc cattagctca ttttatgttag gagttttaaa	3200
tgtgtgtttt tttcccttca agtcttacaa agctaagact ttttgctca ttccttttg	3260
catggttgtc tagggtttct ggacaatgtg ctgttgcatt tttatttcc tagccttgc	3320
aaaatcttcc cttctcaag actttgagca gttagaagtg ctcttttagaa gttgtctgt	3380
ggtgatgtta ctgttagtggc ctcagggaaa ggattgtcca gttacttttag ggggttttg	3440
gtggggtttt tccccctgtg aaaacttact ttgcccctag tctggctgtc gctaggactt	3500
ctgaggagca atgggacatg agtgccttg tatctgcgcc actgcccaca gggaaaggctc	3560
aggaaccaggc acctggaggc caggatagcc aagccctggg tgagcgagag gctggagaac	3620
acaggagctc acccagggtc gtcggccaa catggccac tgtgaacaga cttcagtcct	3680
ctgtttttgt ttcataagcc gttgagacat ctgtggact tggcttaggc cctgctggga	3740
catcccacgt gtgatccctt tcactccatc aggacaccag gactgtcctt aggaaaatgt	3800
ccttgagatg gcagcaggag tcataatttc tgtgtgtgtg tttcgaaag cctgtgtgc	3860
ctgccccatc acaaagaccc agtgcattt gtccttcctg ttctgtgtc actccagaac	3920
ctcagcagat ctgagccacc gctgcccagt gtgagaggcg gccacttca tggcagetta	3980
tcaggcgcag ggccccagac agttccctag ccggccctag agccggccct gggccaaatga	4040
tggagggcgg ccaccagccc agggccctgcc catccagaag ggactccca gggccctgggg	4100
gaggagaccc ttggaaaaagt cctctttcc cagtcctga ttctggatct gagattctca	4160
gatcacagggc ccctgtgtc caggccgagg ctggccacc ctcagggaga tccagagact	4220
catgccccatg gccatccatg cgtggacgct gtgtggagag tccaggatga cgggatcccg	4280
cacaagctcc cttcagtcct tcaggcgtgg gccatgtggg tgattttct aaagctggag	4340

aaggaaagaa ttgtgccttg catattactt gagcttaaac tgacaacctg gatgtaaata	4400
ggagccttgc tactggttta ttataataaag ttctatgtga ttttttaaaa aaaaaaaaaaa	4460
aaaaaaaaaaa aaaaaaaaaaa aaaaatgagc ggcc	4494

<210> 1173
 <211> 1902
 <212> DNA
 <213> Homo sapiens.--

<220>
 <221> CDS
 <222> (986)..(1426)

<400> 1173	
ttatttcaaa gatgagaaaat atttgtttgg atatctatga ctgtctgttt atactgtaaag	60
gggottaatc agcagctcca tcttttagtt ttagttctaa aggaaaagta gcctaaagtc	120
agtataacta aagggtggaa cgaggtggga caaggtccgg aattgctgct cagtgtatgt	180
tgtgtgcctg ccgctggtgg agctgagact gctcatctca gaaggatggg gatgcttgat	240
ttcctggcca ggttgtccca gcacagtggg gattggccct gttgtatgac gaagacagca	300
catggtggca gagatagata ctaacccatg gactttccaa gggagggaaat aggtctttgg	360
agggtatgca agacaaaggt agacactgga taaagaaccc ggtagtcccc aggtattacc	420
ccatctggc cattactccc acactcagga accagacgtt gtgggtgagg acatgctgtc	480
cctcctgcca agtaataact tccttcccag ccaggatcct gccccaaagta ggaatatagc	540
tctgcattta cagcagctcc tgctcagacc ttgtcaaaac caccctgcag cttaggatta	600
aggagcatgg tcacaggaag gtggggtttc agggcatccc ctcaggaact gcccatctcc	660
ccagaattcc aaaatgaagg tccatatgct ttaggtgtg ctggcatgg tgggcttcac	720
agtaggaaag ggtaagtggg gcccaggggc agggagggag gaaggggtaa ctgagttccag	780
gaagggggtg gagcgtggcc atggataatc gggcttccta ctggcccagg gtatttgaga	840
gtgaccaggat gcctccatcc ctccctctgc ctccccagtt cctgttcccg acatccggac	900
gtgccacttc tgcctctgtag aagacccttc tgttagatgc atttcaggct cagagaagtg	960
taccatcagc agctcatccc tgtgc atg gtg atc acc atc tat tat gat gtc	1012
Met Val Ile Thr Ile Tyr Tyr Asp Val	
1	5

aag gtt cgc ttc atc gtt cga ggc tgt gga cag tac att tcc tac cgc	1060		
Lys Val Arg Phe Ile Val Arg Gly Cys Gly Gln Tyr Ile Ser Tyr Arg			
10	15	20	25

tgc caa gaa aaa cgc aac acc tac ttt gca gag tac tgg tat cag gcc	1108		
Cys Gln Glu Lys Arg Asn Thr Tyr Phe Ala Glu Tyr Trp Tyr Gln Ala			
30	35	40	

cag tgc tgt cag tac gat tat tgc aac tcc tgg tca agc ccc caa ctc	1156
---	------

<212> DNA
 <213> Homo sapiens

<400> 4743

attttatctt	tggaaaaaaa	gataaaatgg	atccaaatag	atccacttgc	ttgttaatta	60
ctttcgta	ctgtgcgtt	tttctgagaa	ctagccta	tgtttctgtt	tctctttatc	120
atcatcaa	atgcagct	cgactacctg	tcatttatcc	cgctctcgga	tatcatgtga	180
tat	ttgtctt	ttaaaaaacac	tgcgtcata	tttgtgtt	ataacttgc	240
tttctcccc	ctaggagaa	tcattataga	ttctaaaaat	atatttccc	ttctctgtgg	300
acttggata	aaacgtagct	tttttctgc	ttggatttat	tttctaaaaa	tcaacaccgt	360
aaacccat	cagatacaac	aaaattgggg	tagttaaaac	catgagg	gaaatgagt	420
ttgtggaaac	attaaaaaaa	attgttatac	ccaaagctga	taatcta	ggagaagact	480
ttgactgg	gtttgagg	gttgaagatg	aatcg	ttttttct	gaagtgg	540
tgaatgaaca	gaacgttt	tctgaaagag	aattggaa	tttttagcatt	cttcagaaat	600
caggcaagcc	tattctagaa	ggggcggcat	tggatgaa	tcttaaaacg	tgtaaaactt	660
ctgat	tttgc	gacac	ctggatgata	aagagctg	gaaatttagag	720
aaactctact	gaaattaaag	aacctaaaaa	tccagc	gac	taataatgt	780
cttcagtaac	tagccacaa	tctctgaggt	taatgctaa	agaagaagaa	gccactaaaa	840
agctgaagca	gagtcaagga	attctaaatg	caatgatc	taagatc	aatgaaacttc	900
aggctttac	tgtatgaa	acacaaattg	taatgtt	cagacattt	tttaggtgc	960
aagg	ccactgtt	tttttatcgc	aattttc	ggaaaat	ctaagt	1020
aagagcaaa	cacagc	ttaactt	tttgc	ttttt	cagg	1080
atgaagtagt	tgaaaagtcc	aatgaaagcc	aattttca	ctt	ttttt	1140
catctat	tttgc	gataatcaa	gaaatcctt	aggagag	actagagat	1200
agctcgcata	catttgc	caacatc	taattc	aaa	acgt	1260
tgaagtcaag	tataaaatgg	gcagagg	gtttc	cac	gac	1320
acaaagaaaa	tttggat	gtt	tttgc	tttgc	tttgc	1380
aagaggtc	tc	aaataaa	gacaga	tttgc	tttgc	1440
tattgaat	tttgc	tttgc	tttgc	tttgc	tttgc	1500
at	tttgc	tttgc	tttgc	tttgc	tttgc	1560
acagcaag	acaag	tttgc	tttgc	tttgc	tttgc	1620
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	1680
aaaatttgc	tca	aaaatttgc	tttgc	tttgc	tttgc	1740
taacagat	tttgc	tttgc	tttgc	tttgc	tttgc	1800
attctact	tttgc	tttgc	tttgc	tttgc	tttgc	1860
taactcat	tttgc	tttgc	tttgc	tttgc	tttgc	1920
aagatc	tttgc	tttgc	tttgc	tttgc	tttgc	1980
atgtggacat	tttgc	tttgc	tttgc	tttgc	tttgc	2040
aggagtt	tttgc	tttgc	tttgc	tttgc	tttgc	2100
tcactgat	tttgc	tttgc	tttgc	tttgc	tttgc	2160
atcaat	tttgc	tttgc	tttgc	tttgc	tttgc	2220
atcaat	tttgc	tttgc	tttgc	tttgc	tttgc	2280
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	2303

<210> 4744
 <211> 4458
 <212> DNA
 <213> Homo sapiens

<400> 4744

tttgc	tact	cg	gattc	cagg	agg	gaaatgaa	ggcg	gggg	gggc	ccag	ctct	tg	agcc	cgag	60
ctgt	gtct	ct	cat	tc	ct	gt	ct	at	gt	cc	tct	cc	tct	cc	120
cctgt	gag	gc	ca	ccat	ccat	at	gggg	ccat	tgt	cagg	gt	cc	tttgc	cc	180
atcc	accat	at	gat	ccat	at	ttgt	cc	tgt	cc	ccc	cc	cc	cc	cc	240
ttac	ctgt	tc	gac	ttc	gg	ag	ctgt	at	ctgt	gtgg	tt	cc	tttgc	tt	300
gac	at	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	360
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	420						
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	480						
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	540						

acagattgaa	gtctttcccc	cgttcaggct	gatgccagg	aaggtgacac	tgcttatcgg	600
ggccacatg	caggcacct	ccgagggcgg	cccccagct	cagtccaa	tcctttctc	660
catcagcaat	gagagcgtt	cgctggtgag	cgctgctgg	ctggta	aggcgcct	720
cgggAACGGC	actgtgtctg	ggctcgtgca	gcaagtggat	gcagagacc	gcaagggtgg	780
catcatctct	caggacctcg	tgccaggtgga	ggtgctgctg	ctaaggccg	tgaggatccg	840
cgccccccatc	atgcggatga	ggacggcac	ccagatgcc	atctatgtca	ccggcatcac	900
caaccaccag	aacccttct	ccttggcaa	tgccgtgca	ggcgtgac	tccactggtc	960
tgtcacaag	cgggacgtcc	tggacctccg	agggccgcac	cacgaggcgt	cgatccgact	1020
cccgtcacag	tacaacttt	ccatgaacgt	gctggcccg	gtaaaaggcc	ggaccgggct	1080
gaggggcgtt	gtcaaggctg	tgccagggcac	atcgggccg	ctgtatggcc	tgcccagaga	1140
actctcgat	gagatccaag	tccaggtgtt	tgagaagctg	cagctgtca	accctgaaat	1200
agaagcagaa	caaataattaa	tgtcgc	ccatata	aagctgcaga	caaacaggga	1260
tgtgcagcc	tctctgagct	accgcgtct	gatggacc	gaaaagg	ttcagttgtca	1320
tgttgcgtt	aaaggcttc	tagcatcagg	gtctatgatc	gggacatcca	ccatcgaa	1380
gattgcacaa	gagccctt	gggccaacca	aaccatcatt	gttgcgtt	aaaggatcccc	1440
tgttccctac	ctgaggg	tttccat	ccatgagccc	tgcctgcac	acccagaaca	1500
ggtggccgtg	ccttggaa	tgaccgtgac	ttcactgtc	cacttccac	acaactctgg	1560
agatgtcttc	catgctcaca	gttccgtcct	caacttgc	actaacagag	acgactttgt	1620
gcagatcg	aggggccca	ccaacaacac	ctgtgttgc	cgcacagt	ca	1680
gacactgctc	cgtgtgtgg	acgcaaagca	ccggggcctc	toggact	tgcccctgc	1740
tgtcctacag	ccatctcc	cagagctgtc	tgccatgc	gtgtgtgg	acgtgtctg	1800
tctggccact	gttctgacca	gcttggaa	cctctcagga	acctggag	cctcggccaa	1860
cagcactc	cacatcgacc	ccaagacggg	tgtggctgt	gccccggcc	tggatecg	1920
gacggttac	tatgagg	ctggccac	gggacat	aaggagg	ttggcagcgt	1980
ccctcagagg	atcgg	gtcaacttca	ccccatcc	accagctt	aggaggctac	2040
agcctccaaa	gtgattgtt	ccgtggaga	cagaagct	aacctgag	g	2100
ccccacccag	agggaa	gttccag	ccatcc	g	gtgccc	2160
ccagttcaag	ccggccgt	ttgattt	atctcaagat	gttgc	accg	2220
gtttgacact	gctctcg	acttct	ctcaatcaca	atgcac	aggc	2280
gcagcggaa	cac	tgaa	agac	agctctgg	gtcag	2340
cagccacttc	tccac	agag	ccatcc	acc	catca	2400
cgaccaggct	gaaat	tttcc	tgacca	acc	gtc	2460
tgtcccgag	tttcc	acttgg	gaaat	gttgc	tttgg	2520
cgcaaaaggag	tttcc	gggt	ccat	tcc	ac	2580
cccccgccgt	ggcag	ccat	tttcc	actt	cc	2640
caaccaagcc	atgg	ccat	tttcc	ccat	cc	2700
tccttatg	ccac	acttgg	tttcc	ccat	cc	2760
gtcttcg	ctgtt	ggac	tttcc	ccat	cc	2820
gccccgggat	tttgc	gttgc	tttcc	ccat	cc	2880
ccactattt	gtgc	tttgc	tttcc	ccat	cc	2940
cagccctccc	tcagg	ggag	tttcc	ccat	cc	3000
cccgaggat	gggt	ccat	tttcc	ccat	cc	3060
ttcacactgg	aaca	tttcc	ccat	tttcc	ccat	3120
tctttcaca	cag	tttcc	ccat	tttcc	ccat	3180
taaatgtgt	ttttt	tttcc	ccat	tttcc	ccat	3240
tttgc	tttcc	ccat	tttcc	ccat	tttcc	3300
tgtctaaat	tttcc	ccat	tttcc	ccat	tttcc	3360
tgtgggtgt	tttcc	ccat	tttcc	ccat	tttcc	3420
tttgggtgg	tttcc	ccat	tttcc	ccat	tttcc	3480
acttctgagg	tttcc	ccat	tttcc	ccat	tttcc	3540
cctcaggaac	tttcc	ccat	tttcc	ccat	tttcc	3600
gaacacagga	tttcc	ccat	tttcc	ccat	tttcc	3660
tcctctgtt	tttcc	ccat	tttcc	ccat	tttcc	3720
gggacatccc	tttcc	ccat	tttcc	ccat	tttcc	3780
atgtcttga	tttcc	ccat	tttcc	ccat	tttcc	3840
tgtcttc	tttcc	ccat	tttcc	ccat	tttcc	3900
gaacccat	tttcc	ccat	tttcc	ccat	tttcc	3960
gtttatcagg	tttcc	ccat	tttcc	ccat	tttcc	4020
aatgatggag	tttcc	ccat	tttcc	ccat	tttcc	4080
tgggggagga	tttcc	ccat	tttcc	ccat	tttcc	4140
tctcagatca	tttcc	ccat	tttcc	ccat	tttcc	4200
agactcatgc	tttcc	ccat	tttcc	ccat	tttcc	4260
tcctcgacaa	tttcc	ccat	tttcc	ccat	tttcc	4320
tggagaaagg	tttcc	ccat	tttcc	ccat	tttcc	4380

atgtaaatag gagccttta actggttat ttaataaaagt tctattggtg attttttaaa 4440
aaaaaaaaac ctccgtccc 4458

<210> 4745
<211> 1790
<212> DNA
<213> *Homo sapiens*

<400> 4745

tttattcaaa	gatgagaaaat	atttgggttgg	atatctatga	ctgtctgttt	atactgttaag	60
gggcctaatac	agcagctcca	tcttttagtt	ttagttctaa	aggaaaagta	gcctaaagtc	120
agtataacta	aagggtggaa	cgaggtggga	caaggtccgg	aattgctgct	cagtgtatgt	180
tgtgtgcctg	ccgctgggt	agctgagact	gctcatctca	gaaggatggg	gtgcttgat	240
ttcctggcca	ggttgtccca	gcacagtggg	gattggccct	gttgtatgac	gaagacagca	300
catggtgcca	gagatagata	ctaaccatg	gactttccaa	gggagggaaat	aggtctttgg	360
agggtatgca	agacaaaagg	agacactgg	taaagaaccc	ggtagtgc	ccc aggtattacc	420
ccatctgggc	cattactccc	acactcagga	accagacgtt	gtgggtgagg	acatgtgtc	480
cctcctgcca	agtaataact	tccttccca	ccaggatct	gccccaaagta	ggaatata	540
tctgcattta	cagcagetc	tgctcagacc	ttgtcaaaac	caccctgcag	cttaggatta	600
aggagcatgg	tcacaggaag	gtgggggtt	agggcattcc	ctcaggaact	gcccattctcc	660
ccagaattcc	aaaatgaagg	tcatactgt	tgtaggtgt	ctggctatgg	cgggcttcac	720
agtaggaag	ggtatgtgg	gcccagggg	agggaggag	gaggggtt	ctgatgtccag	780
gaaggggggt	gagcgtggcc	atggataatc	gggttccca	ctggcccagg	gtatttgaga	840
gtgaccctgt	gcctccatcc	ctcttctgc	ctcccccagat	cctttccccc	acatccggac	900
gtgccactc	tgccctgt	aaagccctt	tgttaggtgc	atttcaggct	cagagaagtg	960
taccatcagc	agtcataccc	tgtgcattgt	gatcaccatc	tattatgt	tcaagggttc	1020
cttcatcg	cgaggctgt	gacagttacat	ttcctaccgc	tgccaaagaaa	aacgcacac	1080
ctactttgca	gagtactgt	atcaggccca	gtgctgtc	tacgattatt	gcaactctgt	1140
gtcaagcccc	caactccaga	gtctctgc	ggagccccat	gacagggccc	tgccctgc	1200
tctgtctgac	tcccagattc	agtggttcta	ccaggccctg	aacctctccc	tgccccctccc	1260
caatttccat	gctgggacgg	agctgtatgg	cctggacccc	atggtcacac	tgtccctgaa	1320
cctgggctt	tctttgtgt	agtcgtcccg	catgtacttg	ttcctcaata	gttcaggact	1380
tttgggttctt	ccccaggtgt	gactcttgac	acctcaccc	tcgttgaatcc	cacagtgc	1440
atatcttct	gtaacaccct	cagcatctgt	cactgcactc	cagcctgggt	gacacagtgt	1500
gactctgtt	ccaaaaaaa	ggtatggaa	tagaattctg	tgcagatgtc	ctgacttggc	1560
aattttgtgt	ccctgcctca	ctgtctccac	caaccccccgc	ctgttctagt	gttggcttc	1620
ctctctgtct	ctcttgc	cttgc	tctggcttcc	tggccccat	ttcacttc	1680
tgagtcttga	caccatctc	cctaggggg	tgtgagagga	gagggaaagg	tctgttctgc	1740
tcagctccat	qtccccatt	ttcctccaca	ataaaaactgg	ggactgggct		1790

<210> 4746
<211> 1481
<212> DNA
<213> *Homo sapiens*

<400> 4746

ggccgcgggg	cctccgtccg	agggtggtaaa	tcaagattcc	ttgtttctcg	agccccgagcc	60
ggggcccccc	ccccaggtgc	ttctggggcc	tcagggggcc	ggccatcatca	aagggggtgc	120
acctccacc	ctcatcaccc	actccacagg	gaccacett	gtccctcacccg	tgaccaataaa	180
aaatgcacac	agccctggcc	tgtccagagg	gagcccccaa	cagccctcggt	cccagcctgg	240
ctctccagcg	cctggccct	ctgcccagat	ggacctggag	caccactgc	agcccccttctt	300
tgggaccccc	acttctctgc	tgaagaagga	accacctggc	tatgaggaag	ccatgagccaa	360
gcagccaaaa	cagcaggaaa	atggttccctc	aagccagcag	atggacgacc	tgtttgcacat	420
tctcattcag	agcggagaaaa	tttcagcaga	tttcaaggag	ccgcacatccc	tgccaggggaa	480
ggagaagcca	tccccgaaga	cagtctgttg	gtccccctcg	gcagccacagc	catcaccccttc	540
tgctgagctc	ccccaggctg	ccccacccctcc	tccaggctca	ccctccctcc	ctggacgcctc	600

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34263

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/11, 15/63, 15/70, 15/82, 15/85; C07K 14/00
 US CL : 536/23.1; 435/320.1, 455, 468, 530/300, 350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 536/23.1; 435/320.1, 455, 468, 530/300, 350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 MEDLINE, EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WAJIMA et al. The cDNA cloning and transient expression of an ovary-specific 17beta-hydroxysteroid dehydrogenase of chickens. Gene. 1999, Vol.233, pages 75-82	1-11, 13-16, and 19-26
A	US 5,175,095 A (MARTINEAU et al) 29 December 1992 (29.12.1992), see especially columns 3-18.	1-11, 13-16, and 19-26
A	Database PubMed, ID No. 2393392, FREUDENSTEIN et al. mRNA of bovine tissue inhibitor of metalloproteinase: sequence and expression in bovine ovarian tissue. Biochem. Biophys. Res. Commun. August 1990. Vol.171. No. 1. pages 250-256, see Abstract.	1-11, 13-16, and 19-26
A,P	Database PubMed, ID No. 10919256, HENNEBOLD et al. Ovary-selective genes I: the generation and characterization of an ovary-selective complementary deoxyribonucleic acid library. Endocrinology. August 2000. Vol.141. No.8. pages 2725-2734, see Abstract.	1-11, 13-16, and 19-26
A	Database PubMed, ID No. 2760883, BEIL et al. Synthesis of polypeptides by the cervix of the baboon (Papio anubis). J. Reprod. Fertil. July 1989. Vol.86. No.2. pages 535-544, see Abstract.	1-11, 13-16, and 19-26
A,P	Database PubMed, ID No. 10830289, HINSHELWOOD et al. A 278 bp region just upstream of the human CYP19 (aromatase) gene mediates ovary-specific expression in transgenic mice. Endocrinology. June 2000. Vol.141. No.6. pages 2050-2053, see Abstract.	1-11, 13-16, and 19-26

Further documents are listed in the continuation of Box C. See patent family annex.

Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

Date of mailing of the international search report

07 JUN 2001

Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231
 Facsimile No. (703)305-3230

Authorized officer

Michael Woodward

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34263

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This includes 4 invention Groups and 3572 sequence species

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34263

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid. Group I, claims 1-11, 13-16, and 19-26, drawn to nucleic acid molecules, vector molecules and host cells containing said nucleic acids, polypeptides, methods of making said polypeptides and method of detection using said nucleic acids and polypeptides. Group II, claim 12 and 28, drawn to antibodies and method of treatment using composition comprising said antibodies. Group III, claims 17-18, drawn to methods of identifying a binding partner to a polypeptides. Group IV, claim 27, drawn to method of treatment using composition comprising polypeptides.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I encompasses nucleic acids, polypeptides expressed thereby, vectors and host cells containing same, respectively, and methods of making as well as the first method of use of this subject matter. Groups II-V all are directed to different special technical features as summarized as follows: Group II is directed to an antibody and method of treatment using same, which antibody undergoes recognition and binding reactions wherein what is bound is different from what is bound by the compositions of Group I. For example, the polypeptides of Group I do not bind the polypeptides of Group I as the antibody of Group II does. Identification of binding partner and treatment are clearly different special technical features from detection. Group III is directed to the identification of a binding partner of a polypeptide, which is not identified in any of the other Groups and thus clearly contains its own special technical feature. Group IV is directed to treatment, which is a clearly different method than the methods in the other Groups. Thus, in summary, each of Groups I-IV are directed to different special technical features and thus support this lack of unity.

Additionally, each of the claims is directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows: The claims include a series of polynucleotides and the polypeptides encoded thereby as represented by the sequences of SEQ ID Nos: 1-1786, and 3573-5358. Each of these polynucleotide sequences encodes a separate polypeptide and thus represent a separate gene. Therefore, each of these genes defines its own special technical feature. In summary, one species is a gene represented by one polynucleotide sequence and one polypeptide sequence encoded thereby.